

Chapter 3

Visualizing and Exploring Data



Data Visualization

- ▶ **Data visualization** - the process of displaying data (often in large quantities) in a meaningful fashion to provide insights that will support better decisions.
 - Data visualization improves decision-making, provides managers with better analysis capabilities that reduce reliance on IT professionals, and improves collaboration and information sharing.

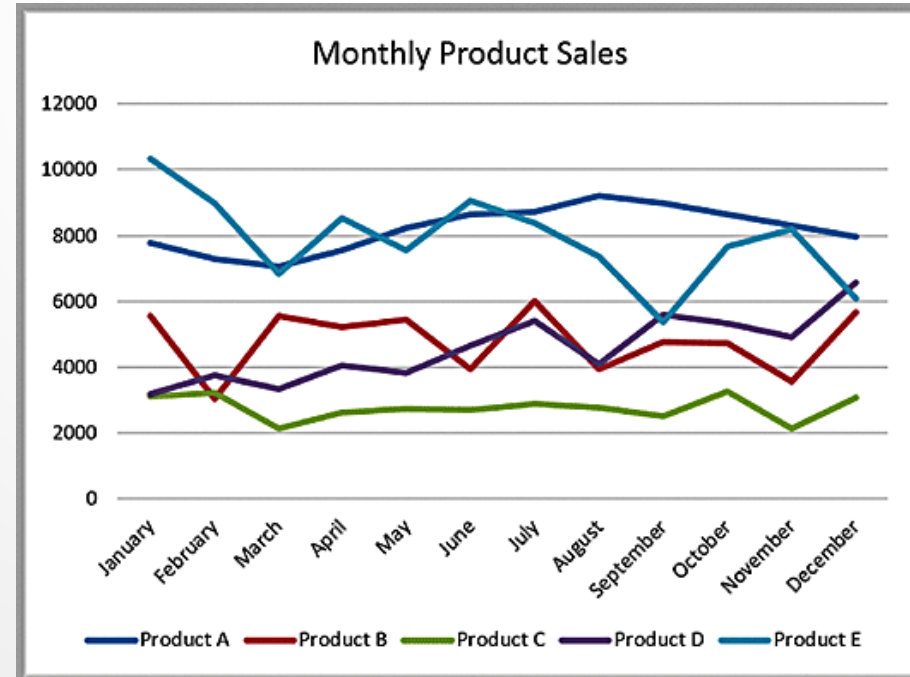
Example 3.1: Tabular vs. Visual Data Analysis

- ▶ Tabular data can be used to determine exactly how many units of a certain product were sold in a particular month, or to compare one month to another.
 - For example, we see that sales of product A dropped in February, specifically by 6.7% (computed as $1 - B3/B2$). Beyond such calculations, however, it is difficult to draw big picture conclusions.

	A	B	C	D	E	F
1	Month	Product A	Product B	Product C	Product D	Product E
2	January	7792	5554	3105	3168	10350
3	February	7268	3024	3228	3751	8965
4	March	7049	5543	2147	3319	6827
5	April	7560	5232	2636	4057	8544
6	May	8233	5450	2726	3837	7535
7	June	8629	3943	2705	4664	9070
8	July	8702	5991	2891	5418	8389
9	August	9215	3920	2782	4085	7367
10	September	8986	4753	2524	5575	5377
11	October	8654	4746	3258	5333	7645
12	November	8315	3566	2144	4924	8173
13	December	7978	5670	3071	6563	6088

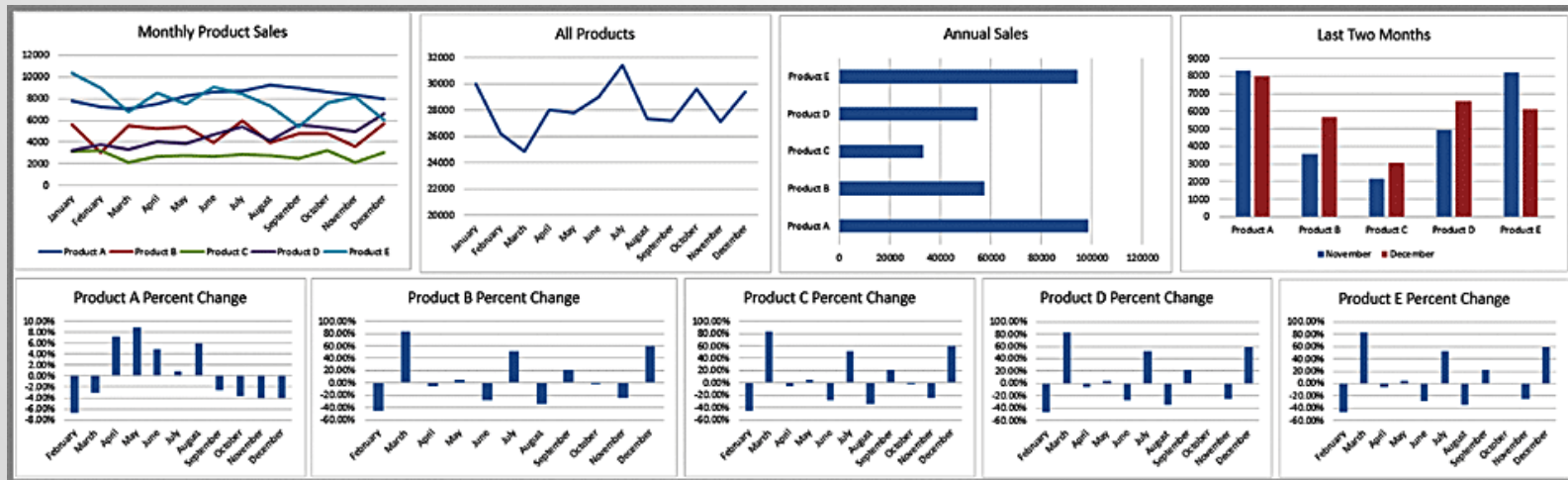
Example 3.1: Tabular vs. Visual Data Analysis

- ▶ A visual chart provides the means to
 - easily compare overall sales of different products (Product C sells the least, for example);
 - identify trends (sales of Product D are increasing), other patterns (sales of Product C is relatively stable while sales of Product B fluctuates more over time), and exceptions (Product E's sales fell considerably in September).



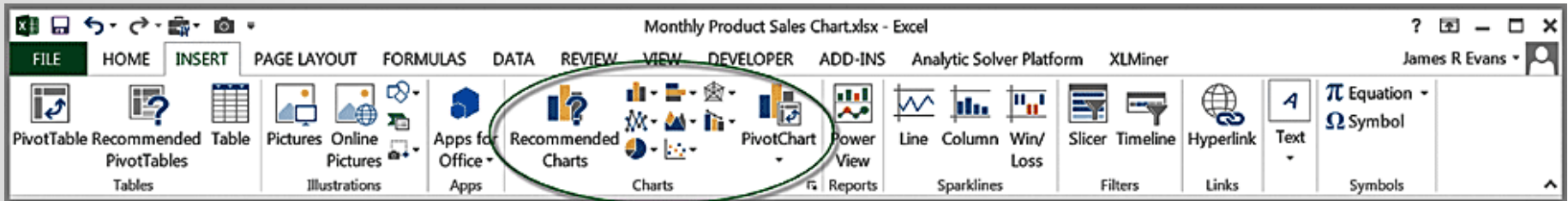
Dashboards

- ▶ A **dashboard** is a visual representation of a set of key business measures. It is derived from the analogy of an automobile's control panel, which displays speed, gasoline level, temperature, and so on.
 - Dashboards provide important summaries of key business information to help manage a business process or function.



Creating Charts in Microsoft Excel

- ▶ Select the *Insert* tab.
- ▶ Highlight the data.
- ▶ Click on chart type, then subtype.



- ▶ Use *Chart Tools* to customize.



Column and Bar Charts

- ▶ Excel distinguishes between vertical and horizontal bar charts, calling the former *column charts* and the latter *bar charts*.
 - A clustered column chart compares values across categories using vertical rectangles;
 - a stacked column chart displays the contribution of each value to the total by stacking the rectangles;
 - a 100% stacked column chart compares the percentage that each value contributes to a total.
- ▶ Column and bar charts are useful for comparing categorical or ordinal data, for illustrating differences between sets of values, and for showing proportions or percentages of a whole.

Example 3.2: Creating a Column Chart

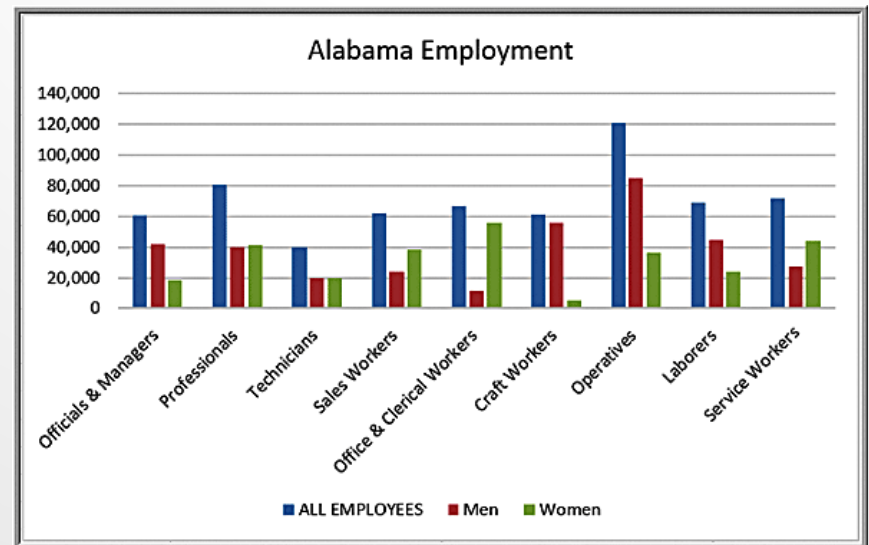
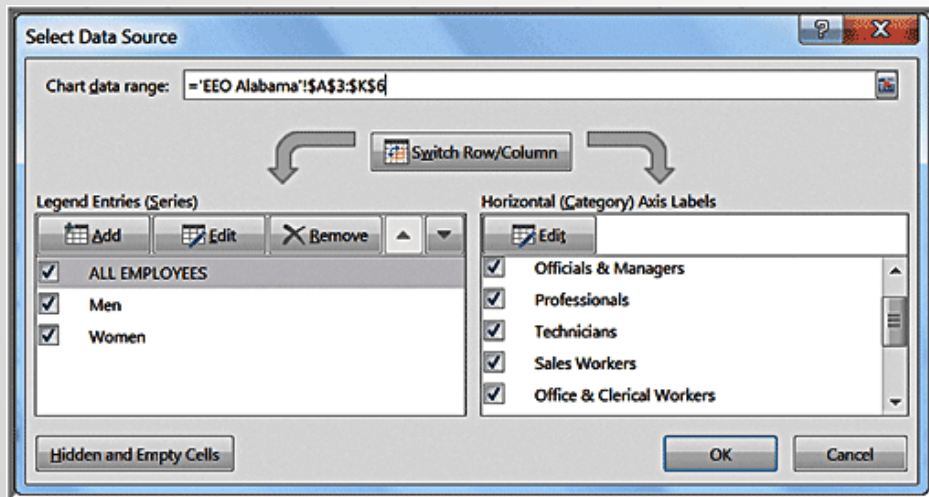
Highlight the range C3:K6, which includes the headings and data for each category. Click on the *Column Chart* button and then on the first chart type in the list (a clustered column chart).

Highlighted Cells

	A	B	C	D	E	F	G	H	I	J	K
1	Equal Employment Opportunity Commission Report - Number Employed in State of Alabama, 2006										
2											
3	Racial/Ethnic Group and Gender	Total Employment	Officials &	Professionals	Technicians	Sales Workers	Office & Clerical	Craft Workers	Operatives	Laborers	Service Workers
4	ALL EMPLOYEES	632,329	60,258	80,733	39,868	62,019	67,014	61,322	120,810	68,752	71,553
5	Men	349,353	41,777	39,792	19,848	23,727	11,293	55,853	84,724	44,736	27,603
6	Women	282,976	18,481	40,941	20,020	38,292	55,721	5,469	36,086	24,016	43,950
7											
8	WHITE	407,545	51,252	67,622	28,830	41,091	44,565	45,742	67,555	26,712	34,176
9	Men	237,516	36,536	34,842	16,004	17,756	7,656	42,699	50,537	17,802	13,684
10	Women	170,029	14,716	32,780	12,826	23,335	36,909	3,043	17,018	8,910	20,492
11											
12	MINORITY	224,784	9,006	13,111	11,038	20,928	22,449	15,580	53,255	42,040	37,377
13	Men	111,837	5,241	4,950	3,844	5,971	3,637	13,154	34,187	26,934	13,919
14	Women	112,947	3,765	8,161	7,194	14,957	18,812	2,426	19,068	15,106	23,458

Example 3.2: Creating a Column Chart

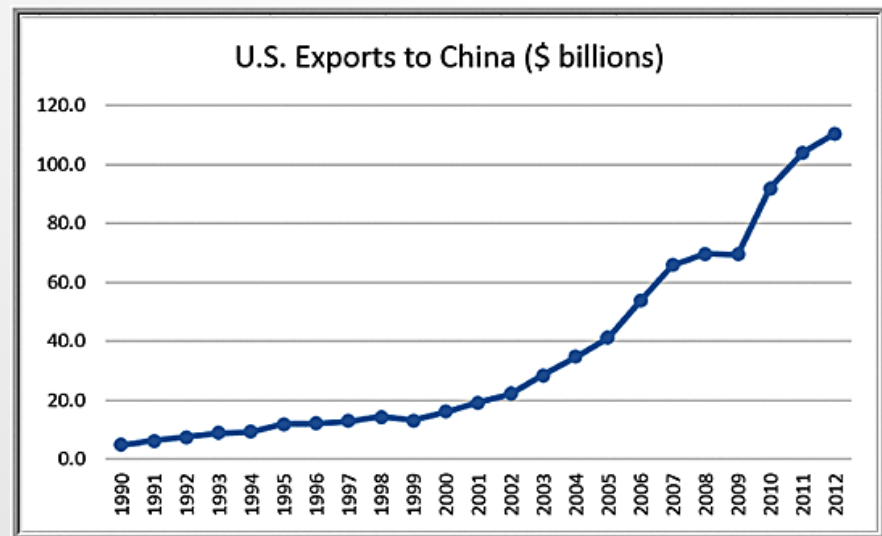
To add a title, click on the first icon in the *Chart Layouts* group. Click on “Chart Title” in the chart and change it to “EEO Employment Report—Alabama.” The names of the data series can be changed by clicking on the *Select Data* button in the *Data* group of the *Design* tab. In the *Select Data Source* dialog (see below), click on “Series1” and then the *Edit* button. Enter the name of the data series, in this case “All Employees.” Change the names of the other data series to “Men” and “Women” in a similar fashion.



Line Charts

- ▶ Line charts provide a useful means for displaying data over time.
 - You may plot multiple data series in line charts; however, they can be difficult to interpret if the magnitude of the data values differs greatly. In that case, it would be advisable to create separate charts for each data series.

Example 3.3: A Line Chart for China Export Data

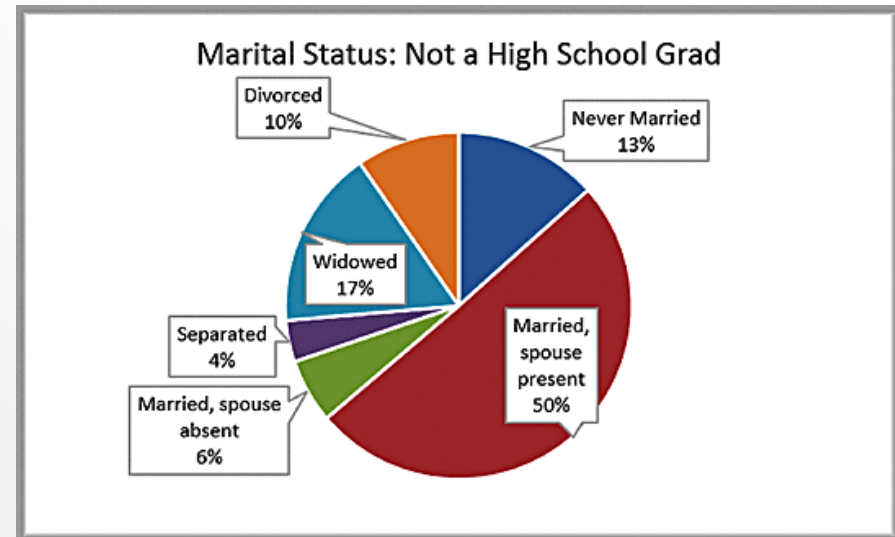


Pie Charts

- ▶ A pie chart displays this by partitioning a circle into pie-shaped areas showing the relative proportion.

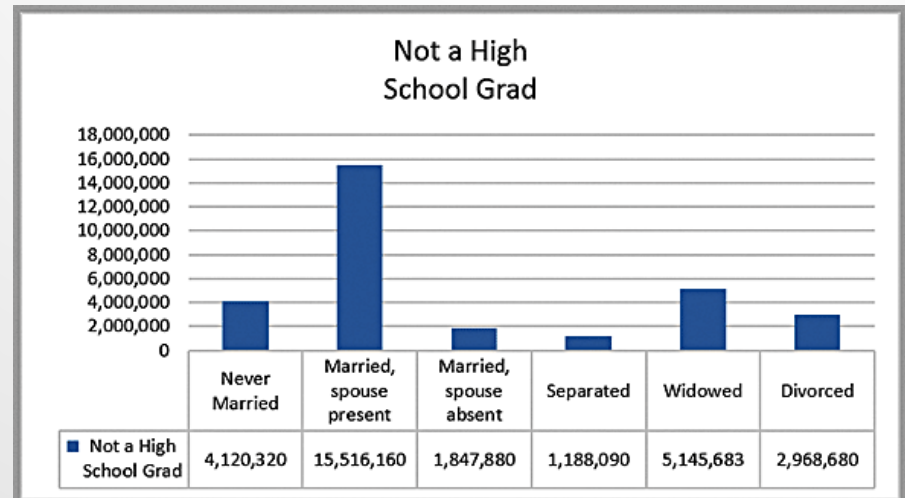
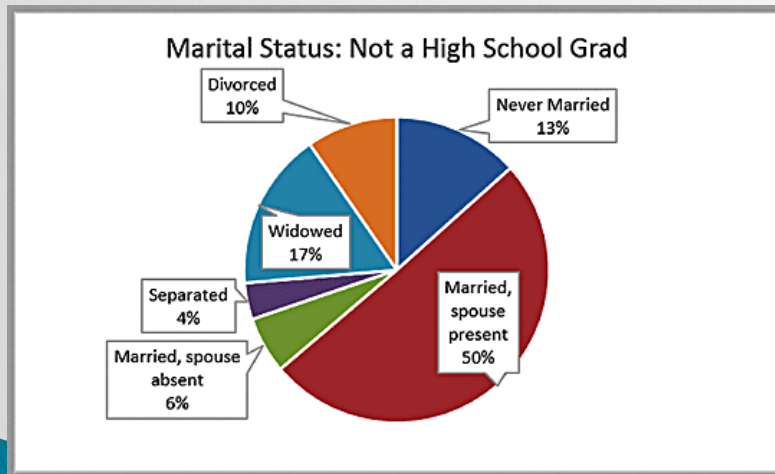
Example 3.4: A Pie Chart for Census Data

	A	B	C	D	E	F	G
1	Census Education Data						
2		Not a High School Grad	High School Graduate	Some College No Degree	Associate's Degree	Bachelor's Degree	Advanced Degree
18	Marital Status						
19	Never Married	4,120,320	7,777,104	4,789,872	1,828,392	5,124,648	2,137,416
20	Married, spouse present	15,516,160	36,382,720	18,084,352	8,346,624	19,154,432	9,523,712
21	Married, spouse absent	1,847,880	2,368,024	1,184,012	465,392	670,712	301,136
22	Separated	1,188,090	1,667,010	842,715	336,165	405,240	165,780
23	Widowed	5,145,683	4,670,488	1,765,010	556,657	977,544	475,195
24	Divorced	2,968,680	7,003,040	3,806,000	1,674,640	2,340,690	1,217,920



Pie Charts

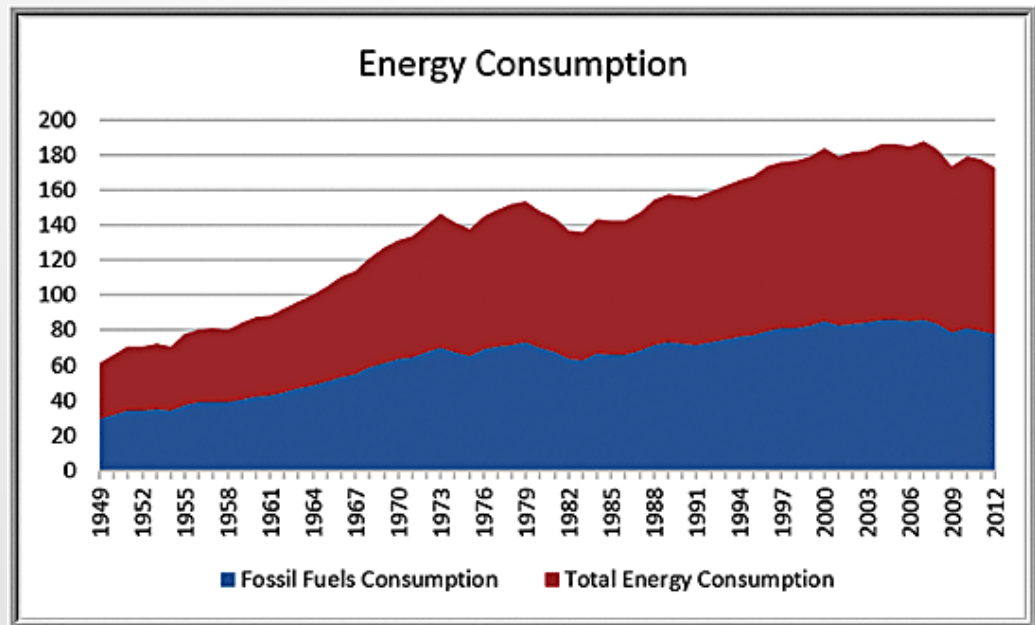
- ▶ Data visualization professionals don't recommend using pie charts. In a pie chart, it is difficult to compare the relative sizes of areas; however, the bars in the column chart can easily be compared to determine relative ratios of the data.
 - If you do use pie charts, restrict them to small numbers of categories, always ensure that the numbers add to 100%, and use labels to display the group names and actual percentages. Avoid three-dimensional (3-D) pie charts—especially those that are rotated—and keep them simple.



Area Charts

- ▶ An area chart combines the features of a pie chart with those of line charts.
 - Area charts present more information than pie or line charts alone but may clutter the observer's mind with too many details if too many data series are used; thus, they should be used with care.

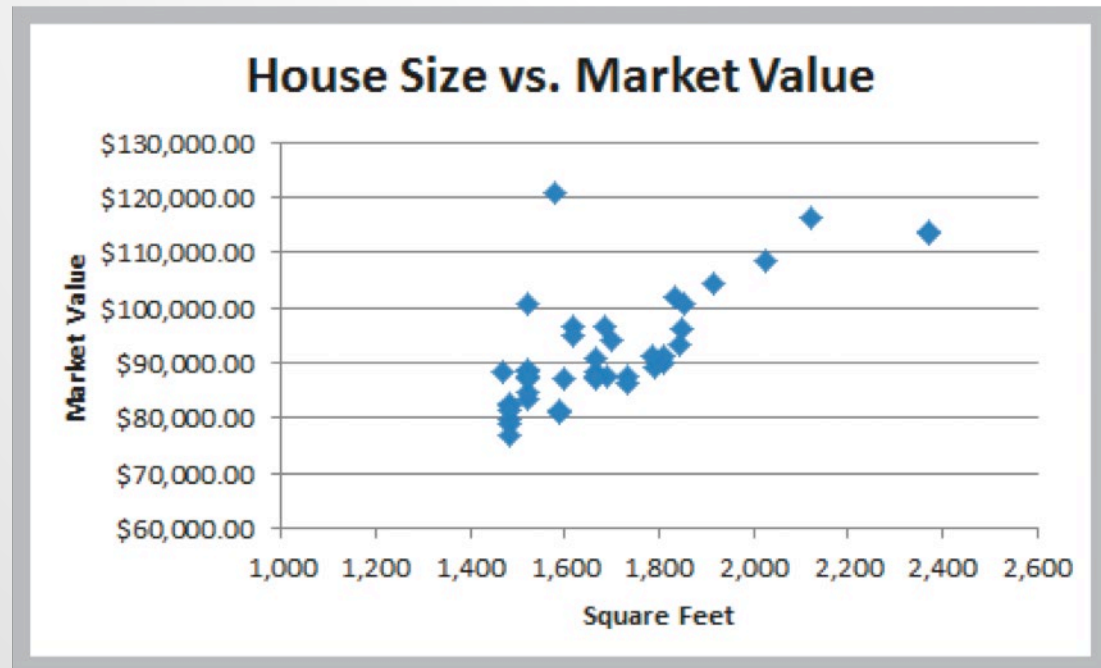
Example 3.5: An Area Chart for Energy Consumption



Scatter Charts

- ▶ Scatter charts show the relationship between two variables. To construct a scatter chart, we need observations that consist of *pairs* of variables.

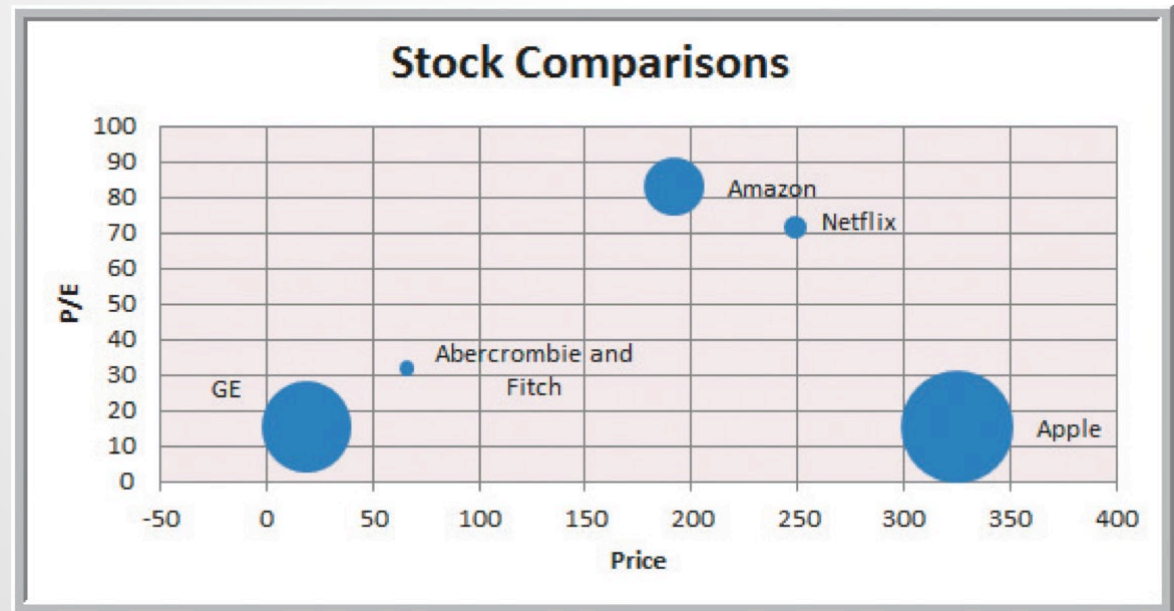
Example 3.6: A Scatter Chart for Real Estate Data



Bubble Charts

- ▶ A bubble chart is a type of scatter chart in which the size of the data marker corresponds to the value of a third variable; consequently, it is a way to plot three variables in two dimensions.

Example 3.7:
A Bubble
Chart for
Stock
Comparisons



Geographic Data

- ▶ Many applications of business analytics involve geographic data. Visualizing geographic data can highlight key data relationships, identify trends, and uncover business opportunities. In addition, it can often help to spot data errors and help end users understand solutions, thus increasing the likelihood of acceptance of decision models.
- ▶ Companies like Nike use geographic data and information systems for visualizing where products are being distributed and how that relates to demographic and sales information. This information is vital to marketing strategies.

Example 3.8: Data Visualization through Conditional Formatting

- ▶ **Data bars** display colored bars that are scaled to the magnitude of the data values (similar to a bar chart) but placed directly within the cells of a range.
 - Highlight the data in each column, click the *Conditional Formatting* button in the *Styles* group within the *Home* tab, select *Data Bars*, and choose the fill option and color.

	A	B	C	D	E	F
1	Month	Product A	Product B	Product C	Product D	Product E
2	January	7792	5554	3105	3168	10350
3	February	7268	3024	3228	3751	8965
4	March	7049	5543	2147	3319	6827
5	April	7560	5232	2636	4057	8544
6	May	8233	5450	2726	3837	7535
7	June	8629	3943	2705	4664	9070
8	July	8702	5991	2891	5418	8389
9	August	9215	3920	2782	4085	7367
10	September	8986	4753	2524	5575	5377
11	October	8654	4746	3258	5333	7645
12	November	8315	3566	2144	4924	8173
13	December	7978	5670	3071	6563	6088

Example 3.8: Data Visualization through Conditional Formatting

- ▶ **Color scales** shade cells based on their numerical value using a color palette.
 - Color-coding of quantitative data is commonly called a **heatmap**.

	A	B	C	D	E	F
1	Month	Product A	Product B	Product C	Product D	Product E
2	January	7792	5554	3105	3168	10350
3	February	7268	3024	3228	3751	8965
4	March	7049	5543	2147	3319	6827
5	April	7560	5232	2636	4057	8544
6	May	8233	5450	2726	3837	7535
7	June	8629	3943	2705	4664	9070
8	July	8702	5991	2891	5418	8389
9	August	9215	3920	2782	4085	7367
10	September	8986	4753	2524	5575	5377
11	October	8654	4746	3258	5333	7645
12	November	8315	3566	2144	4924	8173
13	December	7978	5670	3071	6563	6088

Example 3.8: Data Visualization through Conditional Formatting

- ▶ **Icon sets** provide similar information using various symbols such as arrows or stoplight colors.

	A	B	C	D	E	F
1	Month	Product A	Product B	Product C	Product D	Product E
2	January	↑ 7792	→ 5554	↓ 3105	↓ 3168	↑ 10350
3	February	→ 7268	↓ 3024	↓ 3228	↓ 3751	↑ 8965
4	March	→ 7049	→ 5543	↓ 2147	↓ 3319	→ 6827
5	April	→ 7560	→ 5232	↓ 2636	↓ 4057	↑ 8544
6	May	↑ 8233	→ 5450	↓ 2726	↓ 3837	→ 7535
7	June	↑ 8629	↓ 3943	↓ 2705	↓ 4664	↑ 9070
8	July	↑ 8702	→ 5991	↓ 2891	→ 5418	↑ 8389
9	August	↑ 9215	↓ 3920	↓ 2782	↓ 4085	→ 7367
10	September	↑ 8986	↓ 4753	↓ 2524	→ 5575	→ 5377
11	October	↑ 8654	↓ 4746	↓ 3258	→ 5333	↑ 7645
12	November	↑ 8315	↓ 3566	↓ 2144	→ 4924	↑ 8173
13	December	↑ 7978	→ 5670	↓ 3071	→ 6563	→ 6088

Sparklines

- ▶ **Sparklines** are graphics that summarize a row or column of data in a single cell.
- ▶ Excel has three types of sparklines: line, column, and win/loss.
 - Line sparklines are clearly useful for time-series data
 - Column sparklines are more appropriate for categorical data.
 - Win-loss sparklines are useful for data that move up or down over time.

Example 3.9 Examples of Sparklines

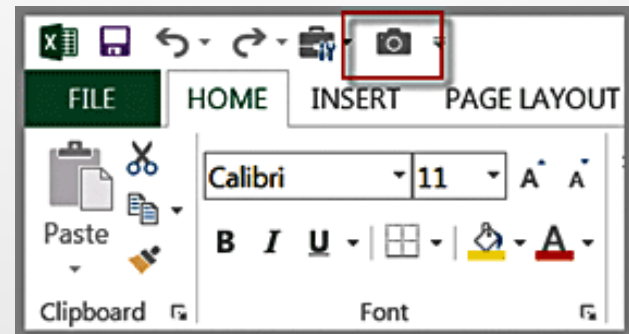
	A	B	C	D	E	F	G
1	Month	Product A	Product B	Product C	Product D	Product E	
2	January	7792	5554	3105	3168	10350	
3	February	7268	3024	3228	3751	8965	
4	March	7049	5543	2147	3319	6827	
5	April	7560	5232	2636	4057	8544	
6	May	8233	5450	2726	3837	7535	
7	June	8629	3943	2705	4664	9070	
8	July	8702	5991	2891	5418	8389	
9	August	9215	3920	2782	4085	7367	
10	September	8986	4753	2524	5575	5377	
11	October	8654	4746	3258	5333	7645	
12	November	8315	3566	2144	4924	8173	
13	December	7978	5670	3071	6563	6088	
14							

	A	B	C	D	E
1	Month	Product A	Percent Change	Product B	Percent Change
2	January	7792		5554	
3	February	7268	-6.72%	3024	-45.55%
4	March	7049	-3.01%	5543	83.30%
5	April	7560	7.25%	5232	-5.61%
6	May	8233	8.90%	5450	4.17%
7	June	8629	4.81%	3943	-27.65%
8	July	8702	0.85%	5991	51.94%
9	August	9215	5.90%	3920	-34.57%
10	September	8986	-2.49%	4753	21.25%
11	October	8654	-3.69%	4746	-0.15%
12	November	8315	-3.92%	3566	-24.86%
13	December	7978	-4.05%	5670	59.00%
14					

- ▶ Generally you need to expand the row or column widths to display them effectively. Notice, however, that the lengths of the bars are not scaled properly to the data; for example, in the first one, products D and E are roughly one-third the value of Product E yet the bars are not scaled correctly. So be careful when using them.

Excel Camera Tool

- ▶ This tool allows you to create live pictures of various ranges from different worksheets that you can place on a single page, size them, and arrange them easily.
- ▶ They are simply linked pictures of the original ranges, and the advantage is that as any data are changed or updated, the camera shots are also.
 - To use the camera too, first add it to the *Quick Access Toolbar* (the set of buttons above the ribbon). From the *File* menu, choose *Options* and then *Quick Access Toolbar*. Choose *Commands*, and then *Commands Not in the Ribbon*. Select *Camera* and add it.



Data Queries: Tables, Sorting, and Filtering

- ▶ Managers often need to sort and filter data.
 - **Filtering** means extracting a set of records having certain characteristics.
- ▶ Excel provides a convenient way of formatting databases to facilitate analysis using sorting and filtering, called *Tables*.

Example 3.10: Creating an Excel Table

- ▶ First, select the range of the data, including headers (a useful shortcut is to select the first cell in the upper left corner, then click *Ctrl+Shift+down arrow*, and then *Ctrl+Shift+right arrow*).
- ▶ Next, click *Table* from the *Tables* group on the *Insert* tab and make sure that the box for *My Table Has Headers* is checked. (You may also just select a cell within the table and then click on *Table* from the *Insert* menu.)
- ▶ The table range will now be formatted and will continue automatically when new data are entered.
- ▶ If you click within a table, the *Table Tools Design* tab will appear in the ribbon, allowing you to do a variety of things, such as change the color scheme, remove duplicates, change the formatting, and so on.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Credit Risk Data											
2												
3	Loan Purp	Checkin	Savin	Months Customer	Months Employ	Gen	Marital Stat	Age	Housi	Years	J	Credit R
4	Small Appliance	\$0	\$739	13	12	M	Single	23	Own	3	Unskilled	Low
5	Furniture	\$0	\$1,230	25	0	M	Divorced	32	Own	1	Skilled	High
6	New Car	\$0	\$389	19	119	M	Single	38	Own	4	Management	High
7	Furniture	\$638	\$347	13	14	M	Single	36	Own	2	Unskilled	High
8	Education	\$963	\$4,754	40	45	M	Single	31	Rent	3	Skilled	Low
9	Furniture	\$2,827	\$0	11	13	M	Married	25	Own	1	Skilled	Low
10	New Car	\$0	\$229	13	16	M	Married	26	Own	3	Unskilled	Low
11	Business	\$0	\$533	14	2	M	Single	27	Own	1	Unskilled	Low
12	Small Appliance	\$6,509	\$493	37	9	M	Single	25	Own	2	Skilled	High
13	Small Appliance	\$966	\$0	25	4	F	Divorced	43	Own	1	Skilled	High
14	Business	\$0	\$989	49	0	M	Single	32	Rent	2	Management	High

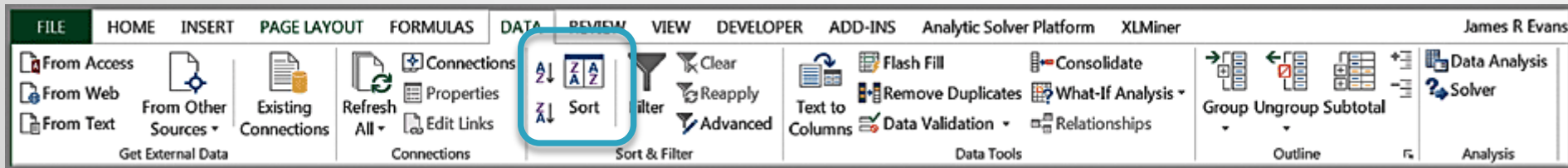
Example 3.11: Table-Based Calculations

- Suppose that in the *Credit Risk Data* table, we wish to calculate the total amount of savings in column C. We could, of course, simply use the function =SUM(C4:C428). However, with a table, we could use the formula =SUM(*Table1*[Savings]). The table name, *Table1*, can be found (and changed) in the *Properties* group of the *Table Tools Design* tab. Note that *Savings* is the name of the header in column C. One of the advantages of doing this is that if we add new records to the table, the calculation will be updated automatically,

	A	B	C	D	E	F	G	H	I	J	K	L
1	Credit Risk Data											
2												
3	Loan Purpo	Checkin	Savin	Months Customer	Months Employ	Gen	Marital Stat	Age	Housi	Years	J	Credit Ri
4	Small Appliance	\$0	\$739	13	12	M	Single	23	Own	3	Unskilled	Low
5	Furniture	\$0	\$1,230	25	0	M	Divorced	32	Own	1	Skilled	High
6	New Car	\$0	\$389	19	119	M	Single	38	Own	4	Management	High
7	Furniture	\$638	\$347	13	14	M	Single	36	Own	2	Unskilled	High
8	Education	\$963	\$4,754	40	45	M	Single	31	Rent	3	Skilled	Low
9	Furniture	\$2,827	\$0	11	13	M	Married	25	Own	1	Skilled	Low
10	New Car	\$0	\$229	13	16	M	Married	26	Own	3	Unskilled	Low
11	Business	\$0	\$533	14	2	M	Single	27	Own	1	Unskilled	Low
12	Small Appliance	\$6,509	\$493	37	9	M	Single	25	Own	2	Skilled	High
13	Small Appliance	\$966	\$0	25	4	F	Divorced	43	Own	1	Skilled	High
14	Business	\$0	\$989	49	0	M	Single	32	Rent	2	Management	High

Sorting Data in Excel

- ▶ The sort buttons in Excel can be found under the *Data* tab in the *Sort & Filter* group. Select a single cell in the column you want to sort on and click the “AZ down arrow” button to sort from smallest to largest or the “AZ up arrow” button to sort from largest to smallest. You may also click the *Sort* button to specify criteria for more advanced sorting capabilities.



Example 3.12 Sorting Data in the Purchase Orders Database

- Suppose we wish to sort the data by supplier. Click on any cell in column A of the data (but not the header cell A3) and then the “AZ down” button in the Data tab. Excel will select the entire range of the data and sort by name of supplier in column A.

	A	B	C	D	E	F	G	H	I	J
1	Purchase Orders									
2										
3	Supplier	Order No.	Item No.	Item Description	Item Cost	Quantity	Cost per order	A/P Terms (Months)	Order Date	Arrival Date
4	Alum Sheeting	Aug11002	1243	Airframe fasteners	\$ 4.25	10,000	\$ 42,500.00	30	08/08/11	08/14/11
5	Alum Sheeting	Sep11002	5417	Control Panel	\$ 255.00	406	\$ 103,530.00	30	09/01/11	09/10/11
6	Alum Sheeting	Sep11008	1243	Airframe fasteners	\$ 4.25	9,000	\$ 38,250.00	30	09/05/11	09/12/11
7	Alum Sheeting	Oct11016	1243	Airframe fasteners	\$ 4.25	10,500	\$ 44,625.00	30	10/10/11	10/17/11
8	Alum Sheeting	Oct11022	4224	Bolt-nut package	\$ 3.95	4,500	\$ 17,775.00	30	10/15/11	10/20/11
9	Alum Sheeting	Oct11026	5417	Control Panel	\$ 255.00	500	\$ 127,500.00	30	10/20/11	10/27/11
10	Alum Sheeting	Oct11028	5634	Side Panel	\$ 185.00	150	\$ 27,750.00	30	10/25/11	11/03/11
11	Alum Sheeting	Oct11036	5634	Side Panel	\$ 185.00	140	\$ 25,900.00	30	10/29/11	11/04/11
12	Durrable Products	Aug11008	7258	Pressure Gauge	\$ 90.00	100	\$ 9,000.00	45	08/25/11	08/28/11
13	Durrable Products	Sep11009	7258	Pressure Gauge	\$ 90.00	120	\$ 10,800.00	45	09/05/11	09/09/11
14	Durrable Products	Sep11027	1369	Airframe fasteners	\$ 4.20	15,000	\$ 63,000.00	45	09/25/11	09/30/11
15	Durrable Products	Sep11031	1369	Airframe fasteners	\$ 4.20	14,000	\$ 58,800.00	45	09/27/11	10/03/11

Pareto Analysis

- ▶ An Italian economist, Vilfredo Pareto, observed in 1906 that a large proportion of the wealth in Italy was owned by a small proportion of the people.
- ▶ Similarly, businesses often find that a large proportion of sales come from a small percentage of customers, a large percentage of quality defects stems from just a couple of sources, or a large percentage of inventory value corresponds to a small percentage of items
- ▶ A **Pareto analysis** involves sorting data and calculating cumulative proportions.

Example 3.13: Applying the Pareto Principle

Sort by

	A	B	C	D	E	F	G	H	I
1	Bicycle Inventory								
2									
3	Product Category	Product Name	Purchase Cost	Selling Price	Supplier	Quantity on Hand	Inventory Value	Percentage	Cumulative %
4	Road	Runroad 5000	\$450.95	\$599.99	Run-Up Bikes	5	\$ 2,254.75	11.2%	11.2%
5	Road	Runroad 1000	\$250.95	\$350.99	Run-Up Bikes	8	\$ 2,007.60	10.0%	21.1%
6	Road	Elegant 210	\$281.52	\$394.13	Bicyclist's Choice	7	\$ 1,970.64	9.8%	30.9%
7	Road	Runroad 4000	\$390.95	\$495.99	Run-Up Bikes	5	\$ 1,954.75	9.7%	40.6%
8	Mtn.	Eagle 3	\$350.52	\$490.73	Bike-One	5	\$ 1,752.60	8.7%	49.3%
9	Road	Classic 109	\$207.49	\$290.49	Bicyclist's Choice	7	\$ 1,452.43	7.2%	56.5%
10	Hybrid	Eagle 7	\$150.89	\$211.46	Bike-One	9	\$ 1,358.01	6.7%	63.3%
11	Hybrid	Tea for Two	\$429.02	\$609.00	Simpson's Bike Supply	3	\$ 1,287.06	6.4%	69.7%
12	Mtn.	Bluff Breaker	\$375.00	\$495.00	The Bike Path	3	\$ 1,125.00	5.6%	75.2%
13	Mtn.	Eagle 2	\$401.11	\$561.54	Bike-One	2	\$ 802.22	4.0%	79.2%
14	Leisure	Breeze LE	\$109.95	\$149.95	The Bike Path	5	\$ 549.75	2.7%	81.9%
15	Children	Runkiddler 100	\$50.95	\$75.99	Run-Up Bikes	10	\$ 509.50	2.5%	84.5%
16	Mtn.	Jetty Breaker	\$455.95	\$649.95	The Bike Path	1	\$ 455.95	2.3%	86.7%
17	Leisure	Runcool 3000	\$85.95	\$135.99	Run-Up Bikes	5	\$ 429.75	2.1%	88.9%
18	Children	Cooltest 100	\$69.99	\$97.98	Bicyclist's Choice	6	\$ 419.94	2.1%	91.0%
19	Mtn.	Eagle 1	\$410.01	\$574.01	Bike-One	1	\$ 410.01	2.0%	93.0%
20	Children	Green Rider	\$95.47	\$133.66	Simpson's Bike Supply	4	\$ 381.88	1.9%	94.9%
21	Leisure	Breeze	\$89.95	\$130.95	The Bike Path	4	\$ 359.80	1.8%	96.7%
22	Leisure	Blue Moon	\$75.29	\$105.41	Simpson's Bike Supply	4	\$ 301.16	1.5%	98.2%
23	Leisure	Supreme 350	\$50.00	\$70.00	Bicyclist's Choice	3	\$ 150.00	0.7%	98.9%
24	Children	Red Rider	\$15.00	\$25.50	Simpson's Bike Supply	8	\$ 120.00	0.6%	99.5%
25	Leisure	Starlight	\$100.47	\$140.66	Simpson's Bike Supply	1	\$ 100.47	0.5%	100.0%
26	Hybrid	Runblend 2000	\$180.95	\$255.99	Run-Up Bikes	0	\$ -	0.0%	100.0%
27	Road	Twist & Shout	\$490.50	\$635.70	Simpson's Bike Supply	0	\$ -	0.0%	100.0%
28						Total	\$ 20,153.27		

75% of the bicycle inventory value comes from 40% (9/24) of items.

Filtering Data

- ▶ For large data files, finding a particular subset of records that meet certain characteristics by sorting can be tedious.
- ▶ Excel provides two filtering tools:
 - AutoFilter for simple criteria, and
 - Advanced Filter for more complex criteria.

Example 3.14: Filtering Records by Item Description

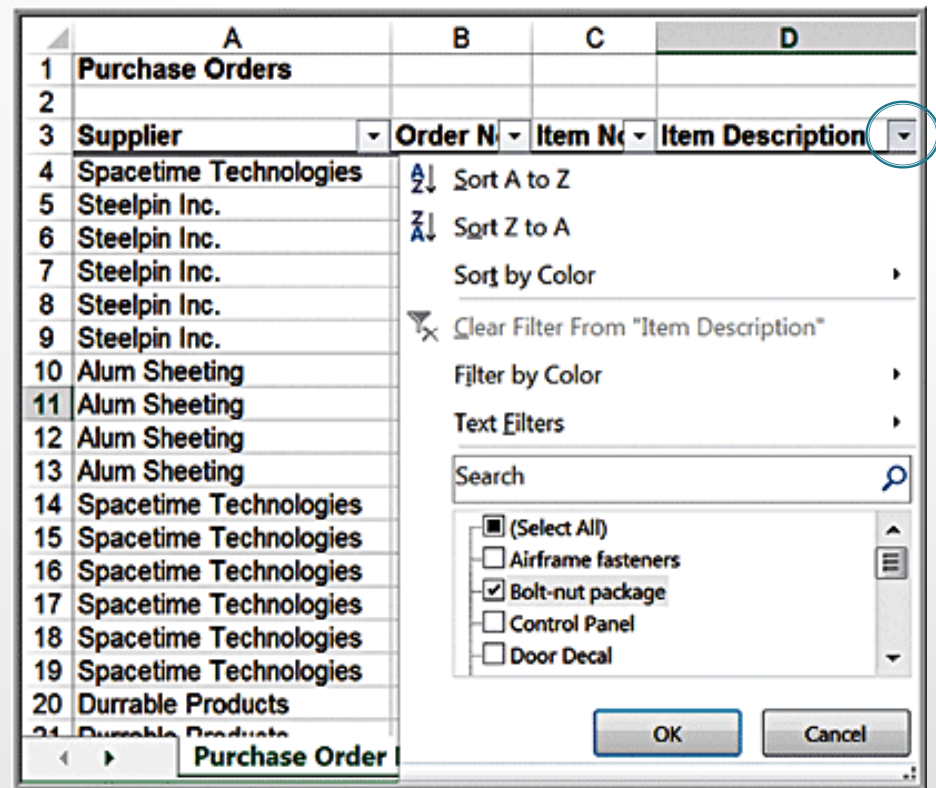
In the *Purchase Orders* database, suppose we are interested in extracting all records corresponding to the item Bolt-nut package.

Select any cell in the database

Data > Sort & Filter > Filter

Click on the dropdown arrow in cell D3.

Select Bolt-nut package to filter out all other items.



Example 3.14: Filter Results

- ▶ The filter tool does not extract the records; it simply hides the records that don't match the criteria. However, you can copy and paste the data to another Excel worksheet, Microsoft Word document, or a Power-Point presentation.
- ▶ To restore the original data file, click on the drop-down arrow again and then click Clear filter from "Item Description."

	A	B	C	D	E	F	G	H	I	J
1	Purchase Orders									
2										
3	Supplier ▼	Order N ▼	Item N ▼	Item Description ▼	Item Co ▼	Quanti ▼	Cost per orde ▼	A/P Terms (Months ▼	Order Dat ▼	Arrival Dat ▼
6	Steelpin Inc.	A0123	4312	Bolt-nut package	\$ 3.75	4,250	\$ 15,937.50	30	08/25/11	09/01/11
9	Steelpin Inc.	A0207	4312	Bolt-nut package	\$ 3.75	4,200	\$ 15,750.00	30	09/01/11	09/10/11
10	Alum Sheeting	A0223	4224	Bolt-nut package	\$ 3.95	4,500	\$ 17,775.00	30	10/15/11	10/20/11
19	Spacetime Technologies	A1222	4111	Bolt-nut package	\$ 3.55	4,200	\$ 14,910.00	25	09/15/11	10/15/11
25	Spacetime Technologies	A1444	4111	Bolt-nut package	\$ 3.55	4,250	\$ 15,087.50	25	09/20/11	10/10/11
26	Spacetime Technologies	A1445	4111	Bolt-nut package	\$ 3.55	4,200	\$ 14,910.00	25	09/25/11	10/25/11
27	Spacetime Technologies	A1449	4111	Bolt-nut package	\$ 3.55	4,600	\$ 16,330.00	25	10/05/11	10/19/11
29	Durrable Products	A1457	4589	Bolt-nut package	\$ 3.50	3,900	\$ 13,650.00	45	10/05/11	10/10/11
35	Spacetime Technologies	A3487	4111	Bolt-nut package	\$ 3.55	4,800	\$ 17,040.00	25	09/05/11	09/20/11
36	Spacetime Technologies	A5689	4111	Bolt-nut package	\$ 3.55	4,585	\$ 16,276.75	25	09/10/11	09/30/11
43	Steelpin Inc.	B0445	4312	Bolt-nut package	\$ 3.75	4,150	\$ 15,562.50	30	09/03/11	09/11/11

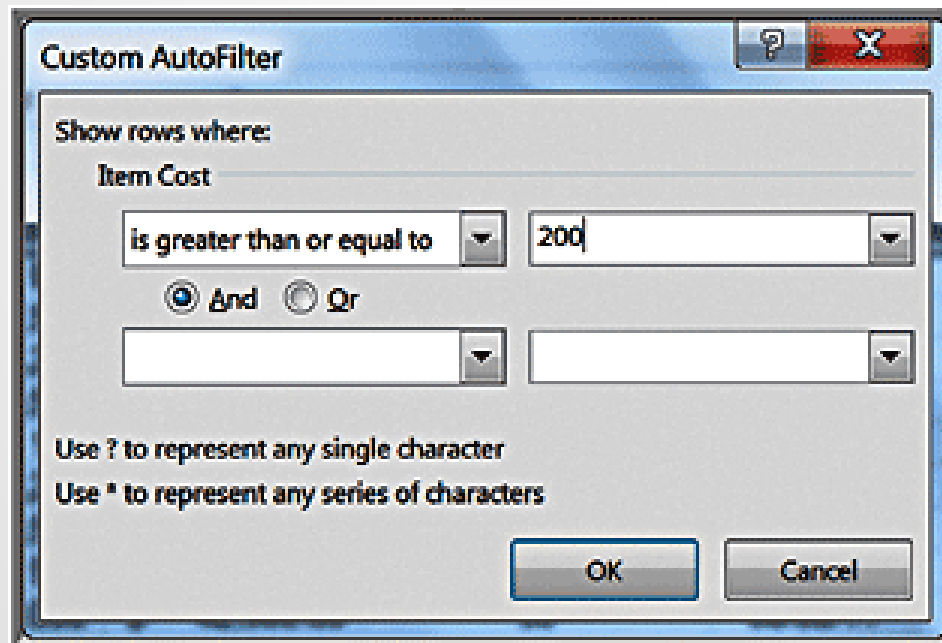
Example 3.15: Filtering Records by Item Cost

- Suppose we wish to identify all records in the *Purchase Orders* database whose item cost is at least \$200. First, click on the drop-down arrow in the Item Cost column and position the cursor over *Numbers Filter*. This displays a list of options. Select *Greater Than Or Equal To . . .* from the list.

	A	B	C	D	E	F	G	H
1	Purchase Orders							
2								
3	Supplier	Order N°	Item N°	Item Description	Item Co.	Quantit	Cost per orde	A/P Terms (Months
4	Spacetime Technologies	A0111	Sort Smallest to Largest			900	\$ 2,700.00	25
5	Steelpin Inc.	A0115	Sort Largest to Smallest			17,500	\$ 19,250.00	30
6	Steelpin Inc.	A0123				4,250	\$ 15,937.50	30
7	Steelpin Inc.	A0204	Sort by Color			16,500	\$ 18,150.00	30
8	Steelpin Inc.	A0205				120	\$ 23,400.00	30
9	Steelpin Inc.	A0207	Clear Filter From "Item Cost"			4,200	\$ 15,750.00	30
10	Alum Sheeting	A0223	Filter by Color			4,500	\$ 17,775.00	30
11	Alum Sheeting	A0433	Number Filters					30
12	Alum Sheeting	A0443	Search					30
13	Alum Sheeting	A0446	<input checked="" type="checkbox"/> (Select All)					30
14	Spacetime Technologies	A0533	<input checked="" type="checkbox"/> \$0.55					30
15	Spacetime Technologies	A0555	<input checked="" type="checkbox"/> \$0.75					25
16	Spacetime Technologies	A0622	<input checked="" type="checkbox"/> \$0.85					25
17	Spacetime Technologies	A0666	<input checked="" type="checkbox"/> \$0.95					25
18	Spacetime Technologies	A0777						25
19	Spacetime Technologies	A1222						25
20	Durrable Products	A1234						45
21	Durrable Products	A1235						45
22	Durrable Products	A1344						45
23	Durrable Products	A1345	9399	Gasket	\$	3.65		45
24	Durrable Products	A1346	9399	Gasket	\$	3.65		45
25	Spacetime Technologies	A1444	4111	Bolt-nut package	\$	3.55		25
26	Spacetime Technologies	A1445	4111	Bolt-nut package	\$	3.55		25
27	Spacetime Technologies	A1449	4111	Bolt-nut package	\$	3.55		25

Example 3.15: Filtering Records by Item Cost

- ▶ The *Custom AutoFilter* dialog allows you to specify up to two specific criteria using “and” and “or” logic. Enter 200 in the box as shown; the tool will display all records having an item cost of \$200 or more.



About the *AutoFilter*

- ▶ *AutoFilter* creates filtering criteria based on the type of data being filtered. If you choose to filter on Order Date or Arrival Date, the *AutoFilter* tools will display a different *Date Filters* menu list for filtering that includes “tomorrow,” “next week,” “year to date,” and so on.
- ▶ *AutoFilter* can be used sequentially to “drill down” into the data.
 - For example, after filtering the results by Bolt-nut package, we could then filter by order date and select all orders processed in September.

Statistical Methods for Summarizing Data

- ▶ **Statistics** is *both the science of uncertainty and the technology of extracting information from data.*
- ▶ A **statistic** is a summary measure of data.
- ▶ **Descriptive statistics** are methods that describe and summarize data.
- ▶ Microsoft Excel supports statistical analysis in two ways:
 1. Statistical functions
 2. *Analysis Toolpak* add-in

Frequency Distributions for Categorical Data

- ▶ A **frequency distribution** is a table that shows the number of observations in each of several nonoverlapping groups.
 - Categorical variables naturally define the groups in a frequency distribution.
- ▶ To construct a frequency distribution, we need only count the number of observations that appear in each category.
 - This can be done using the Excel COUNTIF function.

Example 3.16: Constructing a Frequency Distribution for Items in the *Purchase Orders* Database

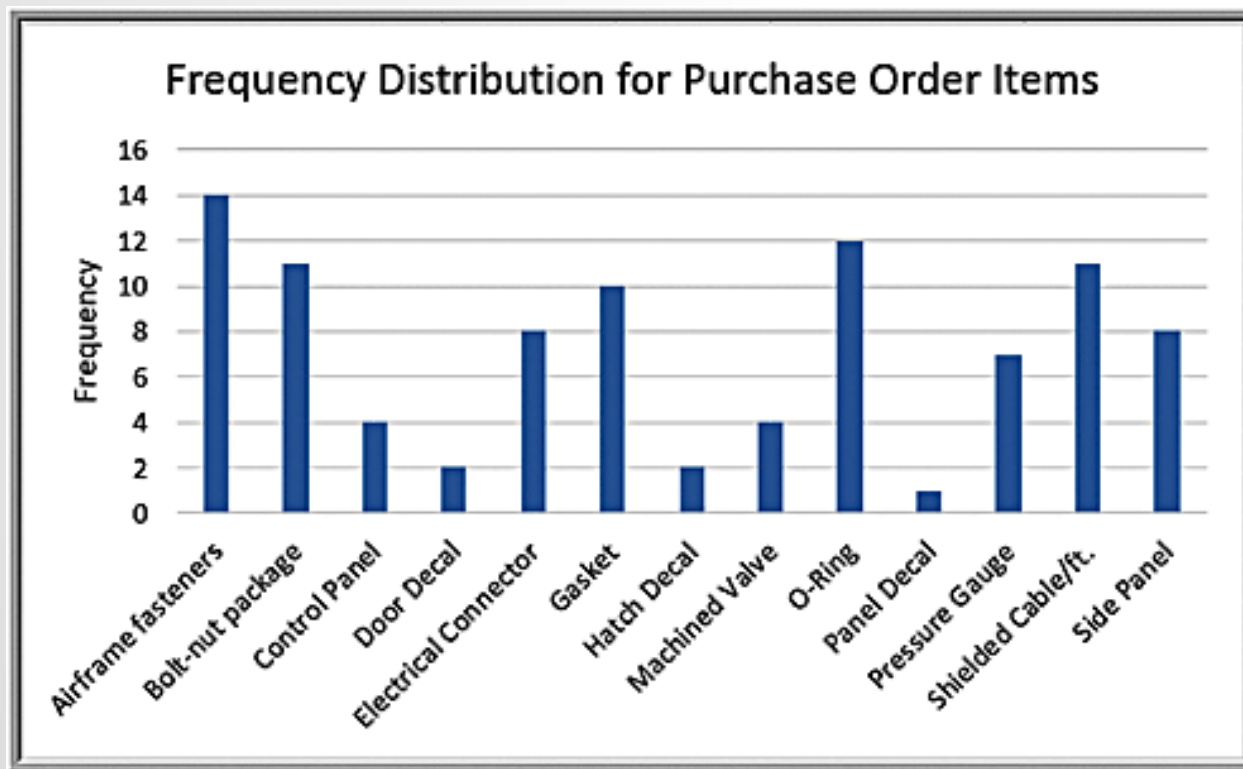
- ▶ List the item names in a column on the spreadsheet.
- ▶ Use the function =COUNTIF(\$D\$4:\$D\$97, *cell_reference*), where *cell_reference* is the cell containing the item name

	A	B
100	Item Description	Frequency
101	Airframe fasteners	=COUNTIF(\$D\$4:\$D\$97,A101)
102	Bolt-nut package	=COUNTIF(\$D\$4:\$D\$97,A102)
103	Control Panel	=COUNTIF(\$D\$4:\$D\$97,A103)
104	Door Decal	=COUNTIF(\$D\$4:\$D\$97,A104)
105	Electrical Connector	=COUNTIF(\$D\$4:\$D\$97,A105)
106	Gasket	=COUNTIF(\$D\$4:\$D\$97,A106)
107	Hatch Decal	=COUNTIF(\$D\$4:\$D\$97,A107)
108	Machined Valve	=COUNTIF(\$D\$4:\$D\$97,A108)
109	O-Ring	=COUNTIF(\$D\$4:\$D\$97,A109)
110	Panel Decal	=COUNTIF(\$D\$4:\$D\$97,A110)
111	Pressure Gauge	=COUNTIF(\$D\$4:\$D\$97,A111)
112	Shielded Cable/ft.	=COUNTIF(\$D\$4:\$D\$97,A112)
113	Side Panel	=COUNTIF(\$D\$4:\$D\$97,A113)

	A	B
100	Item Description	Frequency
101	Airframe fasteners	14
102	Bolt-nut package	11
103	Control Panel	4
104	Door Decal	2
105	Electrical Connector	8
106	Gasket	10
107	Hatch Decal	2
108	Machined Valve	4
109	O-Ring	12
110	Panel Decal	1
111	Pressure Gauge	7
112	Shielded Cable/ft.	11
113	Side Panel	8

Example 3.16: Constructing a Frequency Distribution for Items in the *Purchase Orders* Database

- ▶ Construct a column chart to visualize the frequencies.



Relative Frequency Distributions

- ▶ Relative frequency is the fraction, or proportion, of the total.
- ▶ If a data set has n observations, the relative frequency of category i is:

$$\text{relative frequency of category } i = \frac{\text{frequency of category } i}{n} \quad (3.1)$$

- ▶ We often multiply the relative frequencies by 100 to express them as percentages.
- ▶ A **relative frequency distribution** is a tabular summary of the relative frequencies of all categories.

Example 3.17: Constructing a Relative Frequency Distribution for Items in the *Purchase Orders Database*

- ▶ First, sum the frequencies to find the total number (note that the sum of the frequencies must be the same as the total number of observations, n).
- ▶ Then divide the frequency of each category by this value.

	A	B	C
100	Item Description	Frequency	Relative Frequency
101	Airframe fasteners	14	0.1489
102	Bolt-nut package	11	0.1170
103	Control Panel	4	0.0426
104	Door Decal	2	0.0213
105	Electrical Connector	8	0.0851
106	Gasket	10	0.1064
107	Hatch Decal	2	0.0213
108	Machined Valve	4	0.0426
109	O-Ring	12	0.1277
110	Panel Decal	1	0.0106
111	Pressure Gauge	7	0.0745
112	Shielded Cable/ft.	11	0.1170
113	Side Panel	8	0.0851
114	Total	94	1.0000

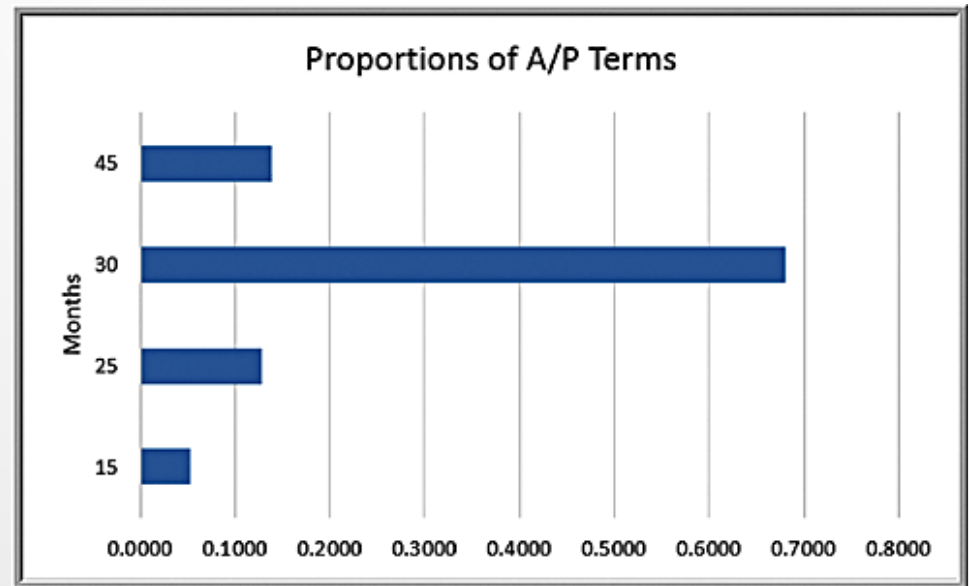
Frequency Distributions for Numerical Data

- ▶ For numerical data that consist of a small number of discrete values, we may construct a frequency distribution similar to the way we did for categorical data; that is, we simply use COUNTIF to count the frequencies of each discrete value.

Example 3.18: Frequency and Relative Frequency Distribution for A/P Terms

- ▶ In the *Purchase Orders* data, the A/P terms are all whole numbers 15, 25, 30, and 45.

	A	B	C
117	A/P Terms	Frequency	Relative Frequency
118	15	5	0.0532
119	25	12	0.1277
120	30	64	0.6809
121	45	13	0.1383
122	Total	94	1.0000

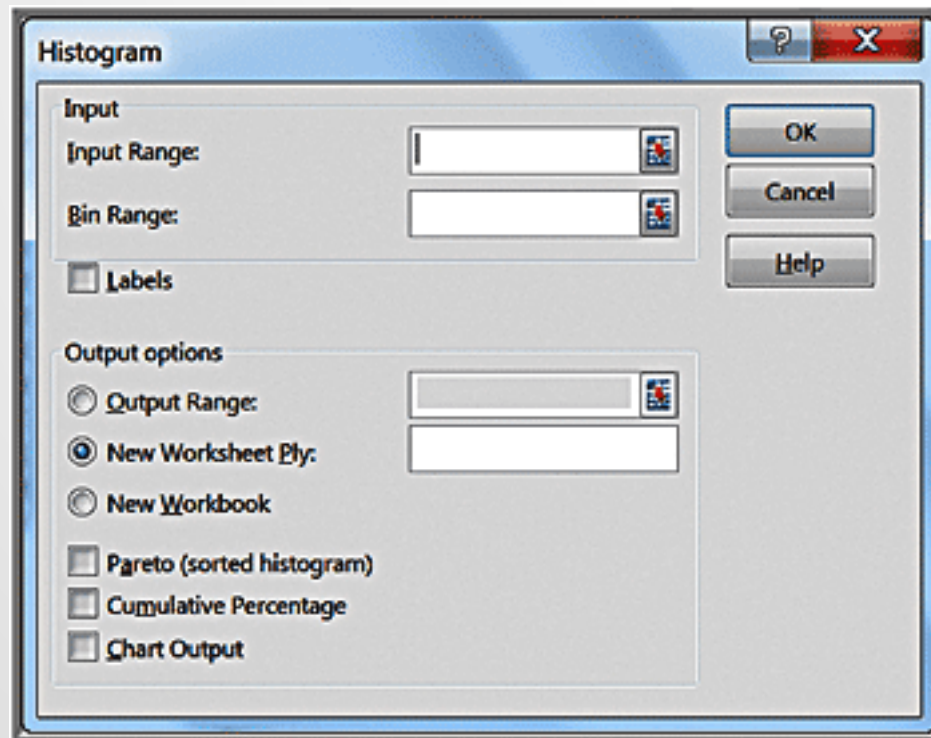


Excel *Histogram* Tool

- ▶ A graphical depiction of a frequency distribution for numerical data in the form of a column chart is called a **histogram**.
- ▶ Frequency distributions and histograms can be created using the *Analysis Toolpak* in Excel.
 - Click the *Data Analysis* tools button in the *Analysis* group under the *Data* tab in the Excel menu bar and select *Histogram* from the list.

Histogram Dialog

- ▶ Specify the *Input Range* corresponding to the data. If you include the column header, then also check the *Labels* box so Excel knows that the range contains a label. The *Bin Range* defines the groups (Excel calls these “bins”) used for the frequency distribution.



Using Bin Ranges

- ▶ If you do not specify a *Bin Range*, Excel will automatically determine bin values for the frequency distribution and histogram, which often results in a rather poor choice.
- ▶ If you have discrete values, set up a column of these values in your spreadsheet for the bin range and specify this range in the *Bin Range* field.

Example 3.19: Using the *Histogram* Tool

- ▶ We will create a frequency distribution and histogram for the *A/P Terms* variable in the *Purchase Orders* database.
- ▶ We defined the bin range below the data in cells H99:H103 as follows:

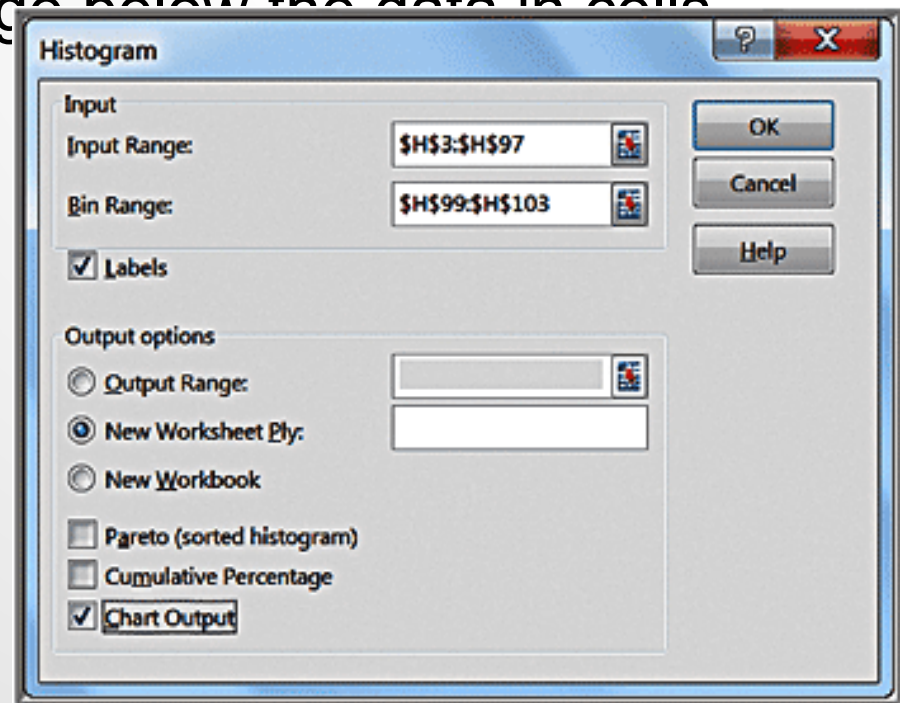
Month

15

25

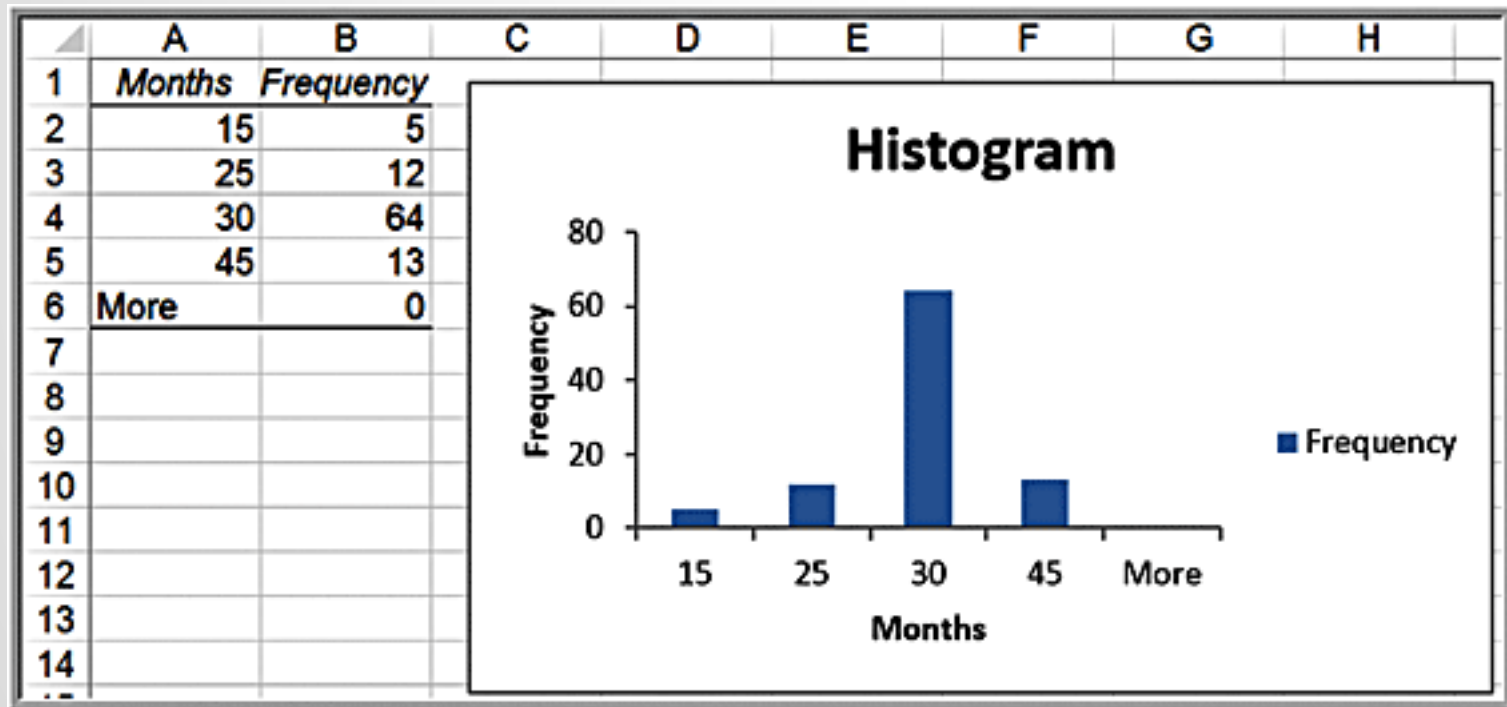
30

45



Example 3.19: Using the *Histogram* Tool

- ▶ *Histogram* tool results:



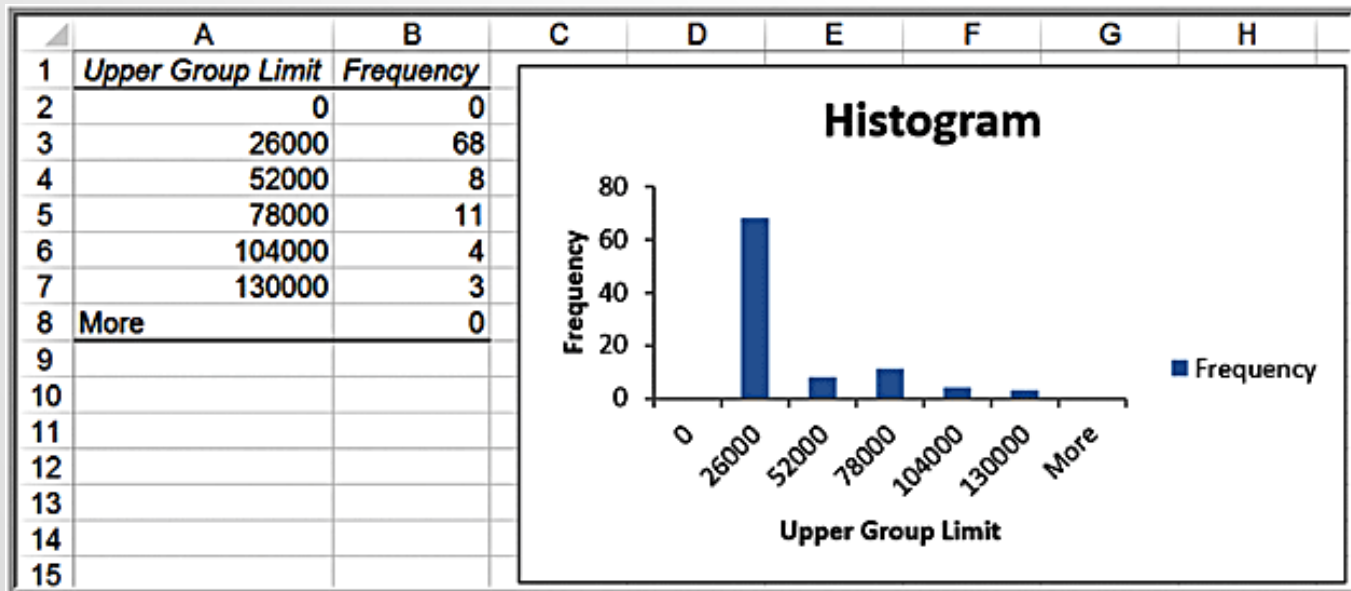
Histograms for Numerical Data

- ▶ For numerical data that have many different discrete values with little repetition or are continuous, a frequency distribution requires that we define by specifying
 1. the number of groups,
 2. the width of each group, and
 3. the upper and lower limits of each group.
- ▶ Choose between 5 to 15 groups, and the range of each should be equal.
- ▶ Choose the lower limit of the first group (LL) as a whole number smaller than the minimum data value and the upper limit of the last group (UL) as a whole number larger than the maximum data value.

$$\text{group width} = \frac{\text{UL} - \text{LL}}{\text{number of groups}} \quad (3.2)$$

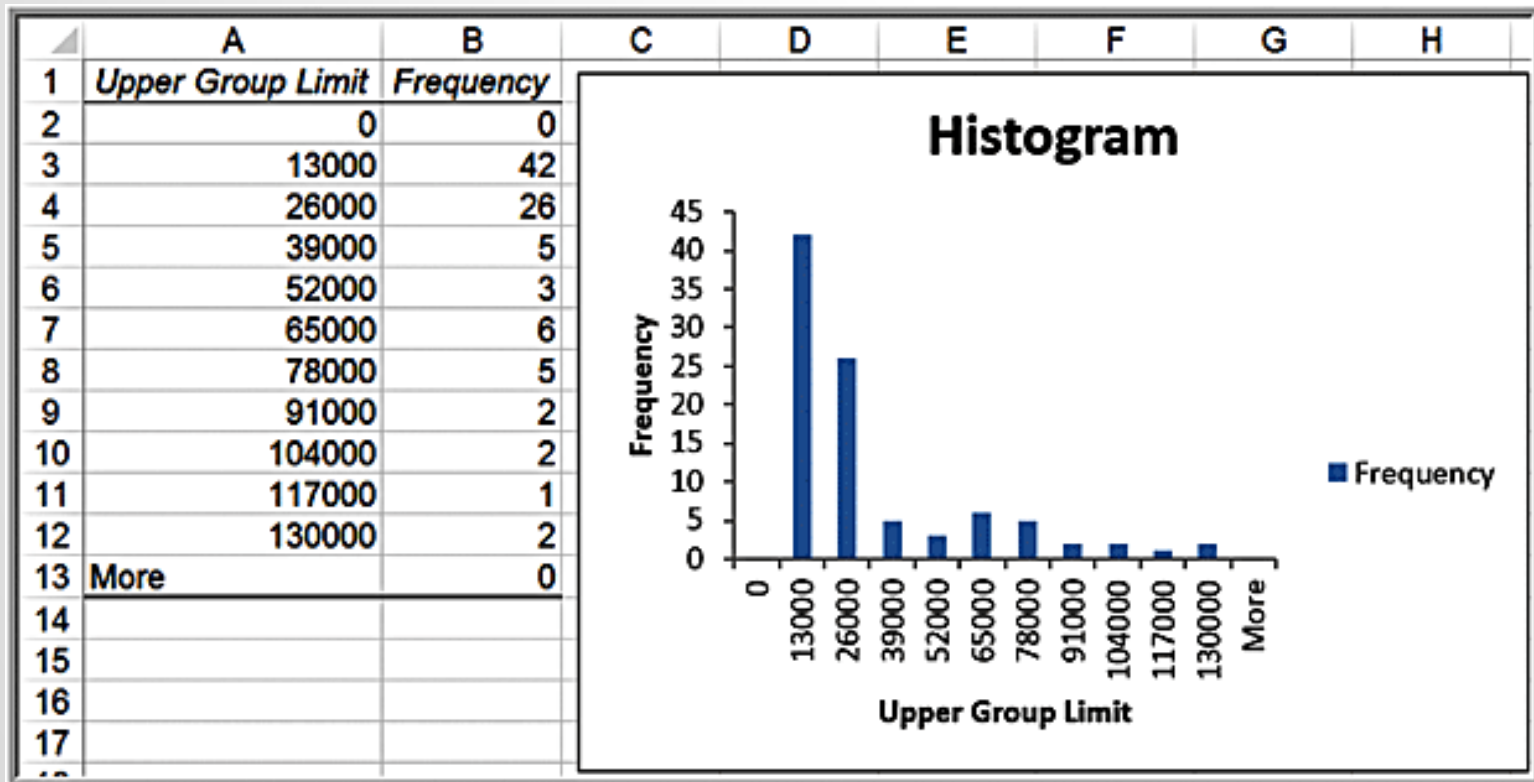
Example 3.20: Constructing a Frequency Distribution and Histogram for Cost per Order

- ▶ The data range from a minimum of \$68.75 to a maximum of \$127,500; set the lower limit of the first group to \$0 and the upper limit of the last group to \$130,000.
- ▶ If we select 5 groups, using equation (3.2) the width of each group is $(\$130,000 - 0) / 5 = \$26,000$



Example 3.20: Constructing a Frequency Distribution and Histogram for Cost per Order

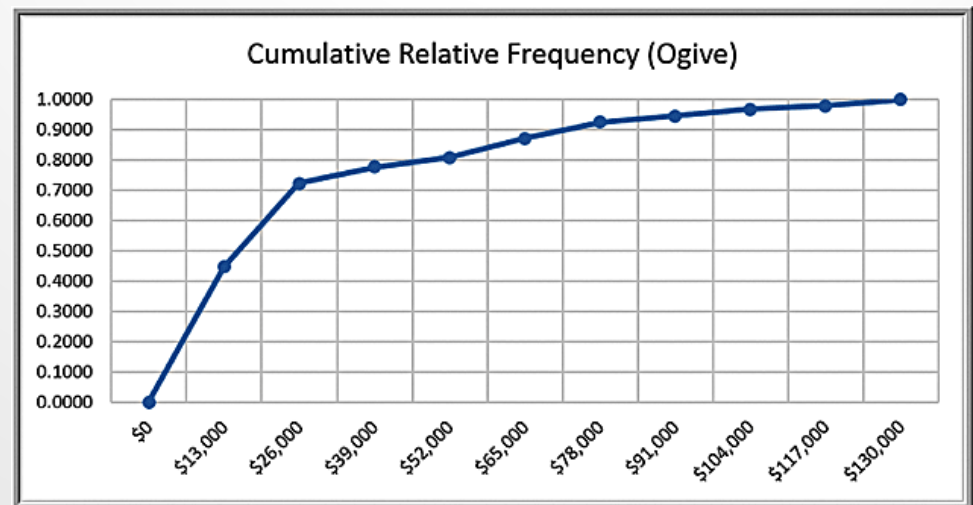
- ▶ Ten-group histogram



Example 3.21 Computing Cumulative Relative Frequencies

- ▶ Set the cumulative relative frequency of the first group equal to its relative frequency. Then add the relative frequency of the next group to the cumulative relative frequency.
- ▶ For, example, the cumulative relative frequency in cell D3 is computed as $=D2+C3 = 0.000 + 0.447 = 0.447$.

	A	B	C	D
1	<i>Upper Group Limit</i>	<i>Frequency</i>	<i>Relative Frequency</i>	<i>Cumulative Relative Frequency</i>
2	0	0	0.0000	0.0000
3	13000	42	0.4468	0.4468
4	26000	26	0.2766	0.7234
5	39000	5	0.0532	0.7766
6	52000	3	0.0319	0.8085
7	65000	6	0.0638	0.8723
8	78000	5	0.0532	0.9255
9	91000	2	0.0213	0.9468
10	104000	2	0.0213	0.9681
11	117000	1	0.0106	0.9787
12	130000	2	0.0213	1.0000
13	More	0	0.0000	1.0000
14	Total	94		



Percentiles

- ▶ The kth percentile is a value at or below which at least k percent of the observations lie. The most common way to compute the kth percentile is to order the data values from smallest to largest and calculate the rank of the kth percentile using the formula:

$$\frac{nk}{100} + 0.5 \quad (3.3)$$

- ▶ Statistical software use different methods that often involve interpolating between ranks instead of rounding, thus producing different results.
 - The Excel function `PERCENTILE.INC(array, k)` computes the kth percentile of data in the range specified in the array field, where k is in the range 0 to 1, inclusive (i.e., including 0 and 1).

Examples 3.22 and 3.23: Computing Percentiles

- ▶ Compute the 90th percentile for *Cost per order* in the *Purchase Orders* data.
 - Rank of k^{th} percentile = $nk/100 + 0.5$
 - $n = 94$; $k = 90$
 - For the 90th percentile, the rank is
$$= 94(90)/100 + 0.5 = 85.1 \text{ (round to 85)}$$
 - Value of the 85th observation = \$74,375
- ▶ Using the Excel function `PERCENTILE.INC(G4:G97,0.9)`, the 90th percentile is \$73,737.50, which is different from using formula (3.3).

Example 3.24 Excel *Rank and Percentile* Tool

Data >

Data Analysis >

Rank and Percentile

90.3rd percentile

= \$74,375

(same result as
manually computing
the 90th percentile)

	A	B	C	D
1	<i>Point</i>	<i>Cost per order</i>	<i>Rank</i>	<i>Percent</i>
2	74	\$127,500.00	1	100.00%
3	62	\$121,000.00	2	98.90%
4	71	\$110,000.00	3	97.80%
5	16	\$103,530.00	4	96.70%
6	73	\$ 96,750.00	5	95.60%
7	1	\$ 82,875.00	6	94.60%
8	67	\$ 81,937.50	7	93.50%
9	82	\$ 77,400.00	8	92.40%
10	54	\$ 76,500.00	9	91.30%
11	80	\$ 74,375.00	10	90.30%
12	68	\$ 72,250.00	11	89.20%
13	20	\$ 65,875.00	12	88.10%
14	65	\$ 64,500.00	13	87.00%
15	28	\$ 63,750.00	14	86.00%

The Excel value of the 90th percentile that was computed in Example 3.23 as \$74,375 is the 90.3rd percentile value.

Quartiles

- ▶ **Quartiles** break the data into four parts.
 - The 25th percentile is called the first quartile, Q1;
 - the 50th percentile is called the second quartile, Q2;
 - the 75th percentile is called the third quartile, Q3; and
 - the 100th percentile is the fourth quartile, Q4.
- ▶ One-fourth of the data fall below the first quartile, one-half are below the second quartile, and three-fourths are below the third quartile.
- ▶ Excel function **QUARTILE.INC(*array*, *quart*)**, where *array* specifies the range of the data and *quart* is a whole number between 1 and 4, designating the desired quartile.

Example 3.25 Computing Quartiles in Excel

- ▶ Compute the Quartiles of the *Cost per Order* data
 - ▶ First quartile: =QUARTILE.INC(G4:G97,1) = \$6,757.81
 - ▶ Second quartile: =QUARTILE.INC(G4:G97,2) = \$15,656.25
 - ▶ Third quartile: =QUARTILE.INC(G4:G97,3) = \$27,593.75
 - ▶ Fourth quartile: =QUARTILE.INC(G4:G97,4) = \$127,500.00

Cross-Tabulations

- ▶ A **cross-tabulation** is a tabular method that displays the number of observations in a data set for different subcategories of two categorical variables.
 - A cross-tabulation table is often called a **contingency table**.
- ▶ The subcategories of the variables must be mutually exclusive and exhaustive, meaning that each observation can be classified into only one subcategory, and, taken together over all subcategories, they must constitute the complete data set.

Example 3.26: Constructing a Cross-Tabulation

- ▶ *Sales Transactions* database

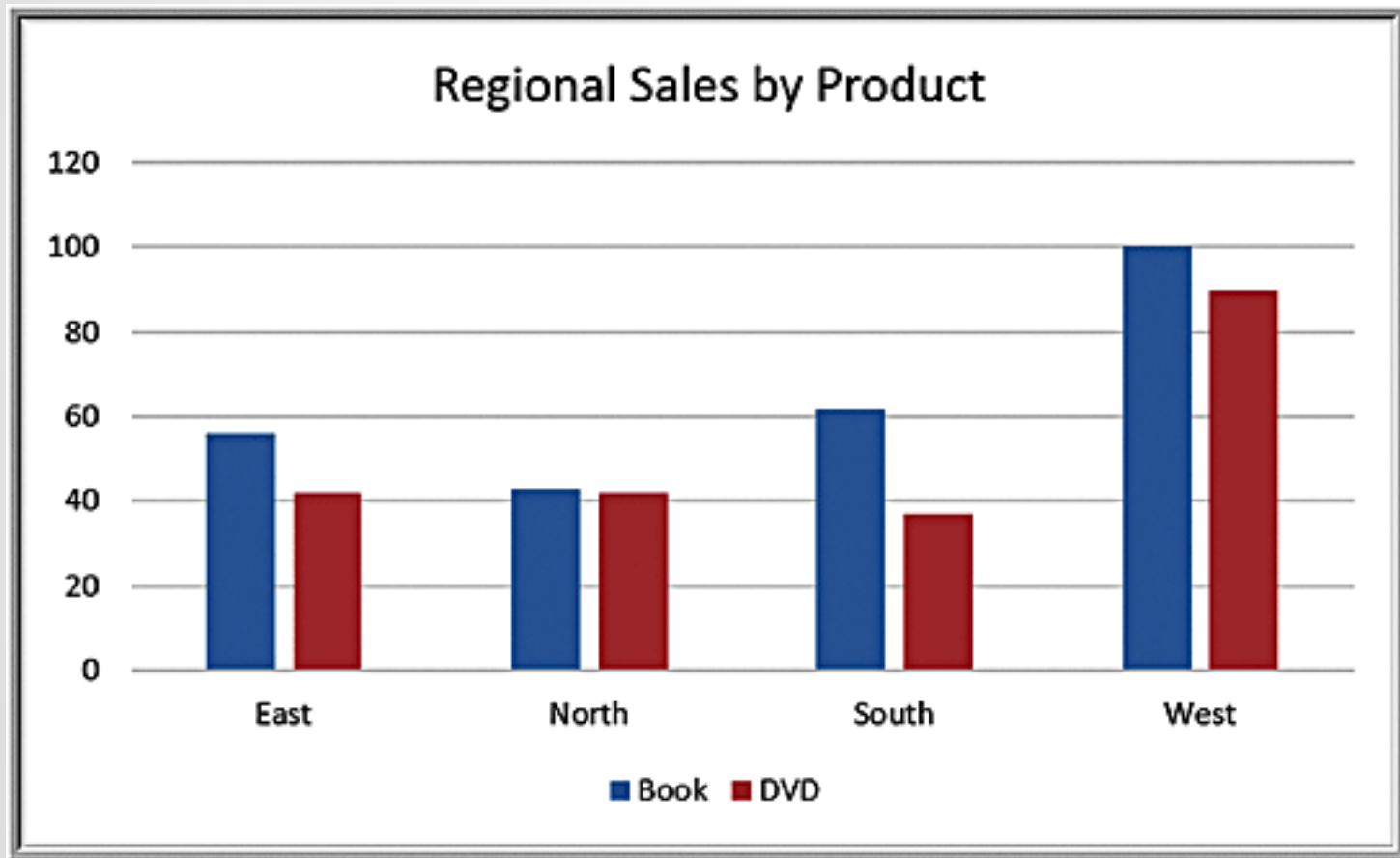
	A	B	C	D	E	F	G	H
1	Sales Transactions: July 14							
2								
3	Cust ID	Region	Payment	Transaction Code	Source	Amount	Product	Time Of Day
4	10001	East	Paypal	93816545	Web	\$20.19	DVD	22:19
5	10002	West	Credit	74083490	Web	\$17.85	DVD	13:27
6	10003	North	Credit	64942368	Web	\$23.98	DVD	14:27
7	10004	West	Paypal	70560957	Email	\$23.51	Book	15:38
8	10005	South	Credit	35208817	Web	\$15.33	Book	15:21
9	10006	West	Paypal	20978903	Email	\$17.30	DVD	13:11
10	10007	East	Credit	80103311	Web	\$177.72	Book	21:59
11	10008	West	Credit	14132683	Web	\$21.76	Book	4:04
12	10009	West	Paypal	40128225	Web	\$15.92	DVD	19:35
13	10010	South	Paypal	49073721	Web	\$23.39	DVD	13:26

- ▶ Count the number (and compute the percentage) of books and DVDs ordered by region.

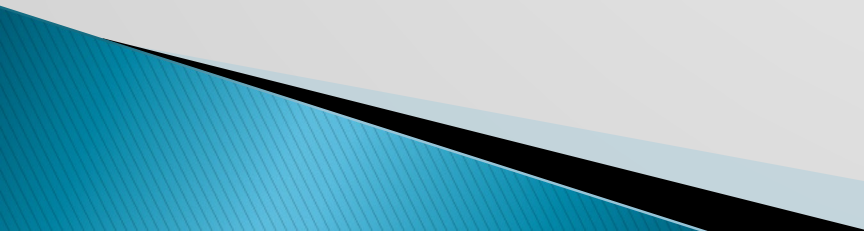
Region	Book	DVD	Total
East	56	42	98
North	43	42	85
South	62	37	99
West	100	90	190
Total	261	211	472

Region	Book	DVD	Total
East	57.1%	42.9%	100.0%
North	50.6%	49.4%	100.0%
South	62.6%	37.4%	100.0%
West	52.6%	47.4%	100.0%

Cross-Tabulation Visualization: Chart of Regional Sales by Product



Exploring Data Using PivotTables

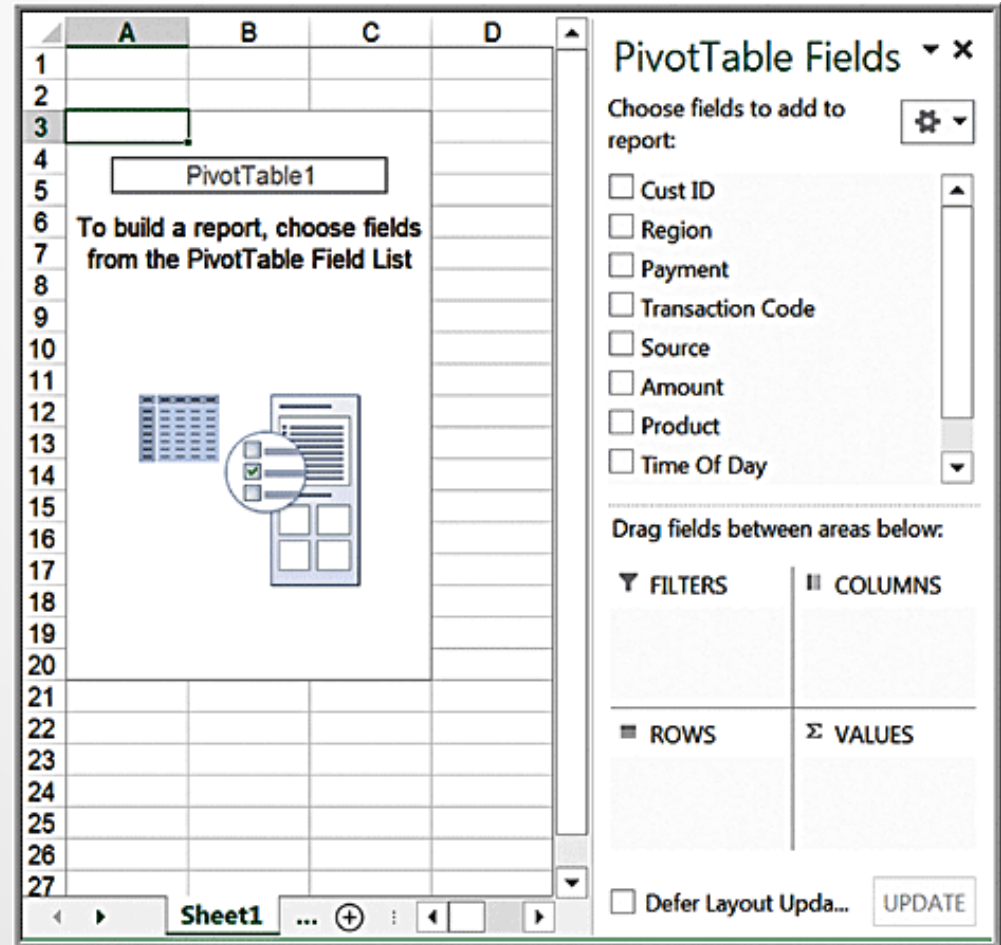
- ▶ Excel provides a powerful tool for distilling a complex data set into meaningful information: **PivotTables.**
 - ▶ PivotTables allows you to create custom summaries and charts of key information in the data.
 - ▶ PivotTables can be used to quickly create cross-tabulations and to drill down into a large set of data in numerous ways.
- 

Constructing PivotTables

Click inside your database

*Insert >
Tables >
PivotTable*

The wizard creates a blank PivotTable as shown.



PivotTable Field List

Select and drag the fields to one of the PivotTable areas:

- ▶ *Report Filter*
- ▶ *Column Labels*
- ▶ *Row Labels*
- ▶ Σ *Values*

The screenshot displays an Excel spreadsheet with a PivotTable named 'PivotTable1' in the center. The PivotTable area contains the text: 'To build a report, choose fields from the PivotTable Field List'. Below this text is a diagram illustrating the PivotTable layout with four quadrants: FILTERS, COLUMNS, ROWS, and VALUES. A blue arrow points from the 'COLUMNS' label in the diagram to the 'COLUMNS' area in the PivotTable Fields task pane.

The PivotTable Fields task pane on the right is titled 'PivotTable Fields' and includes the following elements:

- Choose fields to add to report:** A list of fields with checkboxes: Cust ID, Region, Payment, Transaction Code, Source, Amount, Product, and Time Of Day.
- Drag fields between areas below:** A 2x2 grid of areas: FILTERS (top-left), COLUMNS (top-right), ROWS (bottom-left), and VALUES (bottom-right).
- Defer Layout Update:** A checkbox at the bottom left.
- UPDATE:** A button at the bottom right.

Example 3.27 Creating a PivotTable

Initial PivotTable
for Regional
Sales by Product

The PivotTable
defaults to a sum
of the field in the
Values area.

We seek a count
of the number of
records in each
category.

The screenshot shows an Excel spreadsheet with a PivotTable and the PivotTable Fields task pane. The PivotTable is located in the range A3:D9 and displays the sum of Cust ID for each region and product. The PivotTable Fields task pane is open on the right side of the spreadsheet, showing the fields to be added to the report and the layout of the PivotTable.

Row Labels	Book	DVD	Grand Total
East	572755	428278	1001033
North	441841	429848	871689
South	634963	379724	1014687
West	1024473	919746	1944219
Grand Total	2674032	2157596	4831628

PivotTable Fields

Choose fields to add to report:

- Cust ID
- Region
- Payment
- Transaction Code
- Source
- Amount
- Product
- Time Of Day

Drag fields between areas below:

FILTERS

COLUMNS
Product

ROWS
Region

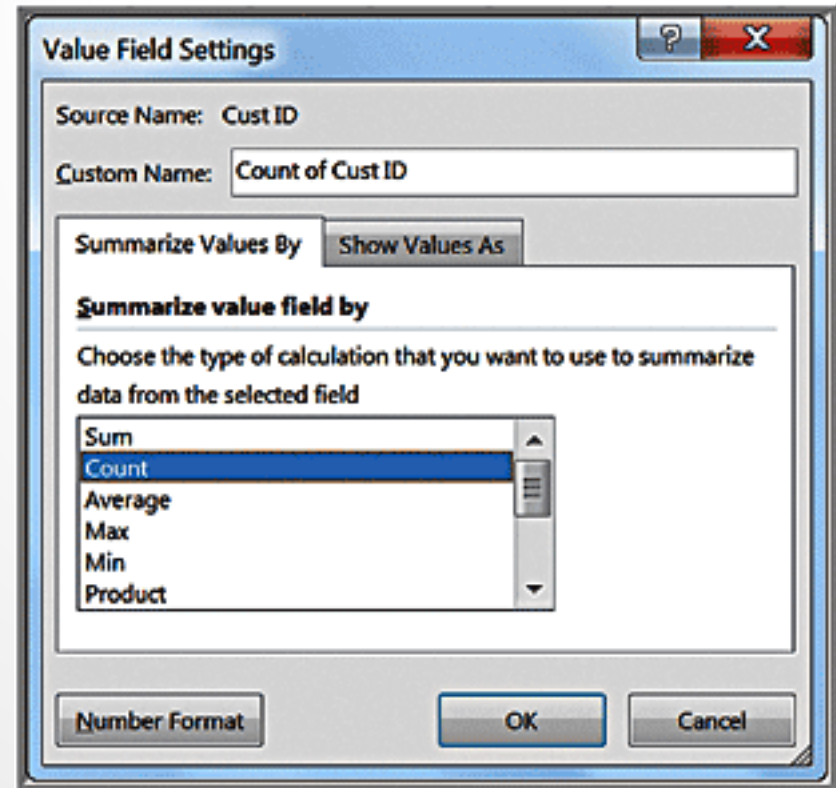
VALUES
Sum of Cust ID

Defer Layout Update UPDATE

Changing Value Field Settings

Active Field > Analyze > Field Settings

- ▶ Change summarization method in *Value Field Settings* dialog box
- ▶ Select *Count*



Final Pivot Table

	A	B	C	D
1				
2				
3	Count of Cust ID	Column Labels ▼		
4	Row Labels ▼	Book	DVD	Grand Total
5	East	56	42	98
6	North	43	42	85
7	South	62	37	99
8	West	100	90	190
9	Grand Total	261	211	472

Modifying PivotTables

- ▶ Uncheck the boxes in the *PivotTable Field List* or drag the field names to different areas.
- ▶ You may easily add multiple variables in the fields to create different views of the data.
 - Example: drag the *Source* field into the *Row Labels* area

The screenshot displays an Excel spreadsheet with a PivotTable and the PivotTable Fields task pane. The PivotTable is located in the range A3:E17 and shows sales data categorized by Region, Source, and Product. The task pane on the right is titled 'PivotTable Fields' and shows the following configuration:

- Choose fields to add to report:** Region (checked), Payment (unchecked), Transaction Code (unchecked), Source (checked), Amount (unchecked), Product (checked), Time Of Day (unchecked).
- MORE TABLES...** (dropdown)
- Drag fields between areas below:**
- FILTERS:** (empty)
- COLUMNS:** Product
- ROWS:** Region, Source
- VALUES:** Count of Cust ID
- Defer Layout Update:** (unchecked)
- UPDATE** (button)

Count of Cust ID	Column Labels	DVD	Grand Total	
East	Book	56	42	98
Email		18	6	24
Web		38	36	74
North		43	42	85
Email		12	13	25
Web		31	29	60
South		62	37	99
Email		20	10	30
Web		42	27	69
West		100	90	190
Email		29	21	50
Web		71	69	140
Grand Total		261	211	472

Example 3.28: Using the PivotTable Report Filter

- ▶ Dragging a field into the Report Filter area in the PivotTable Field list allows you to add a third dimension to your analysis.

The screenshot shows an Excel PivotTable and the PivotTable Fields task pane. The PivotTable is located in the range A3:D9. The PivotTable Fields task pane is on the right, showing the following configuration:

- Choose fields to add to report:** Cust ID, Region, Payment, Product (checked); Transaction Code, Source, Amount, Time Of Day (unchecked).
- Drag fields between areas below:**
- FILTERS:** Payment
- COLUMNS:** Product
- ROWS:** Region
- VALUES:** Count of Cust ID

Count of Cust ID	Column Labels	DVD Grand Total		
Row Labels	Book			
East		56	42	98
North		43	42	85
South		62	37	99
West		100	90	190
Grand Total		261	211	472

Click the drop down arrow in cell B1; choose Credit:

The screenshot shows the updated Excel PivotTable. The PivotTable is located in the range A3:D9. The PivotTable Fields task pane is not visible. The PivotTable is now filtered by Payment = Credit. The PivotTable Fields task pane configuration is:

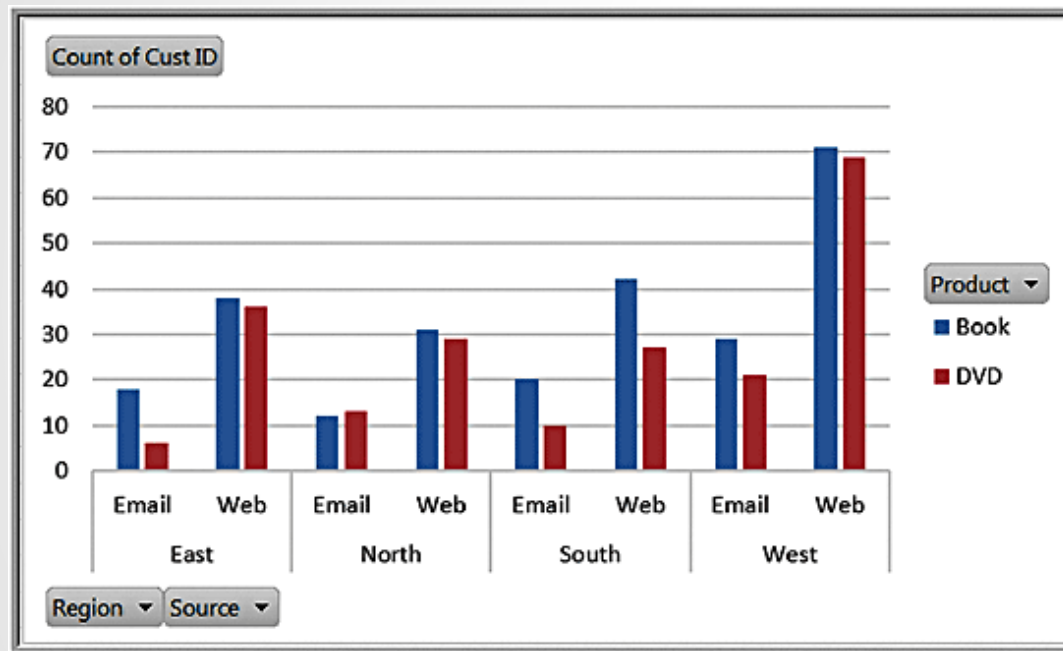
- FILTERS:** Payment
- COLUMNS:** Product
- ROWS:** Region
- VALUES:** Count of Cust ID

Count of Cust ID	Column Labels	DVD Grand Total		
Row Labels	Book			
East		40	34	74
North		21	29	50
South		44	17	61
West		54	60	114
Grand Total		159	140	299

PivotCharts

- ▶ **PivotCharts** visualize data in PivotTables.
- ▶ They can be created in a simple one-click fashion.
 - Select the PivotTable
 - From the analyze tab, click *PivotChart*.
 - Excel will display an *Insert Chart* dialog that allows you to choose the type of chart you wish to display.

Example 3.29: A PivotChart for Sales Data



By clicking on the drop-down buttons, you can easily change the data that are displayed. by filtering the data. Also, by clicking on the chart and selecting the *PivotChart Tools Design* tab, you can switch the rows and columns to display an alternate view of the chart or change the chart type entirely.

Slicers

- ▶ Excel 2010 introduced **slicers** — tools for drilling down to “slice” a PivotTable and display a subset of data.
- ▶ To create a slicer for any of the columns in the database, click on the PivotTable and choose *Insert Slicer* from the *Analyze* tab in the *PivotTable Tools* ribbon.

Example 3.30 Using Slicers

	A	B	C	D	E	F	G	H
1								
2								
3	Count of Cust ID	Column Labels ▼						
4	Row Labels ▼	Book		DVD Grand Total				
5	East		56	42	98			
6	North		43	42	85			
7	South		62	37	99			
8	West		100	90	190			
9	Grand Total		261	211	472			
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								

Insert Slicers [?] [X]

- Cust ID
- Region
- Payment
- Transaction Code
- Source
- Amount
- Product
- Time Of Day

[OK] [Cancel]

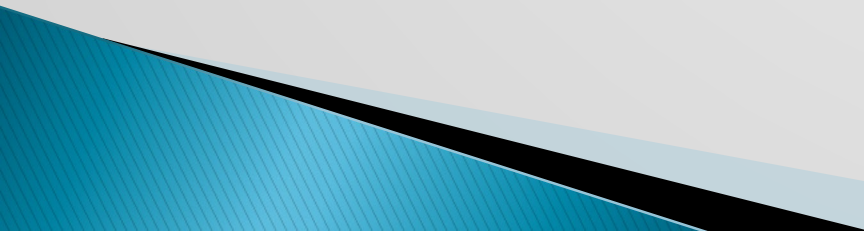
Cross-tabulation
“sliced” by E-
mail

	A	B	C	D	E	F	G
1							
2							
3	Count of Cust ID	Column Labels ▼					
4	Row Labels ▼	Book		DVD Grand Total			
5	East		18	6	24		
6	North		12	13	25		
7	South		20	10	30		
8	West		29	21	50		
9	Grand Total		79	50	129		

Source [Filter Icon]

- Email
- Web

PivotTable Dashboards

- ▶ The camera tool is useful for creating PivotTable-based dashboards.
 - ▶ If you create several different PivotTables and charts, you can easily use the camera tool to take pictures of them and consolidate them onto one worksheet.
 - ▶ In this fashion, you can still make changes to the PivotTables and they will automatically be reflected in the camera shots.
- 

Camera-Based Dashboard Example

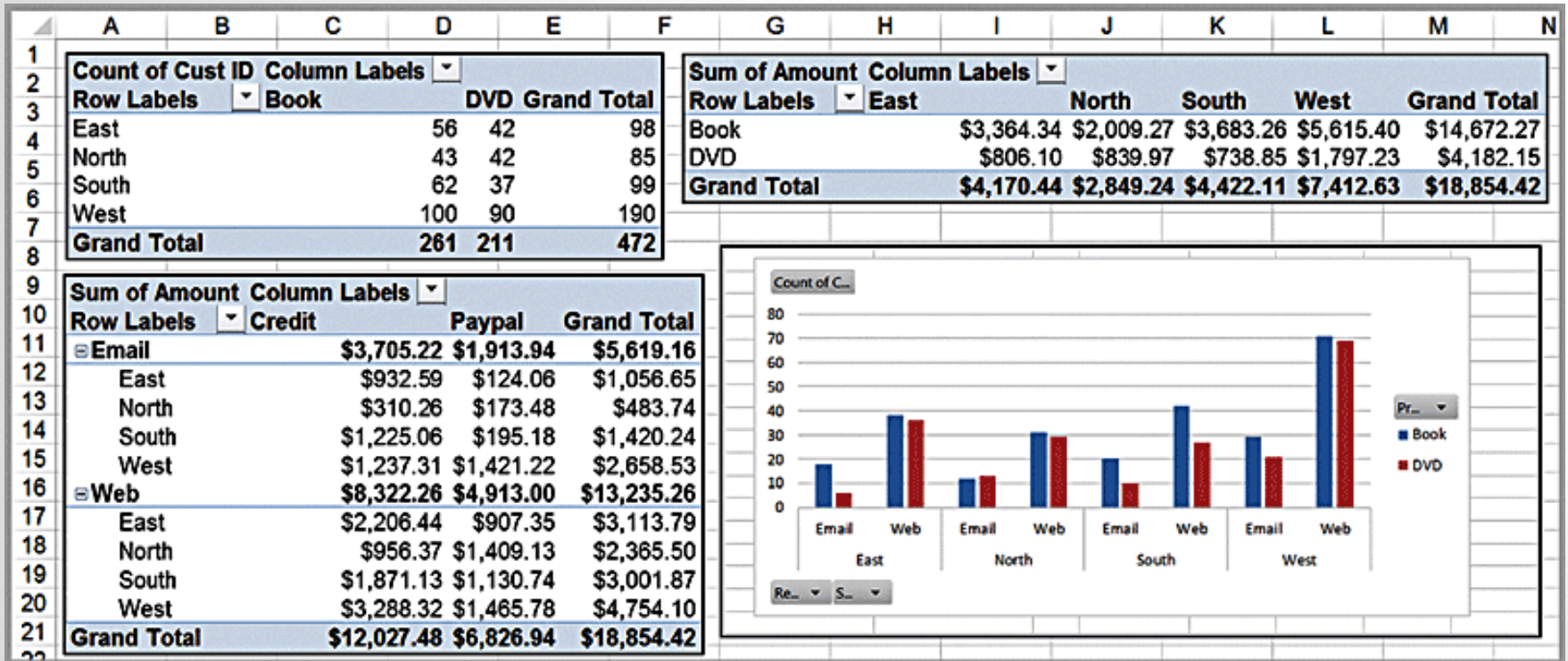


Tableau Introduction

- ▶ Tableau is a visualization tool for data analytics, converting rows and columns of data into a nice graphic to make decisions from.
- ▶ Visit my viz link for some Tableau examples at <https://public.tableau.com/app/profile/albertkalim>

Tableau Instructions

- ▶ 1. Download the free Tableau Public App (Tableau Public 2021.4.4) at: <https://public.tableau.com/en-us/s/>
- ▶ 2. Create a Tableau Public profile (same URL) to share your “viz”. Click on "Sign In" on upper right corner. At bottom of the pop-up screen, click "Create one now for free". Follow the instructions to complete the registration.

Tableau's Big Data

- ▶ Visit <http://www.uky.edu/~akali2/Tableau/> to download big data (in Excel files) to play with in Tableau. Some of the .csv files may/may not work because some Tableau versions may require .xls file types to work.
- ▶ For further trainings and videos, visit <https://www.tableau.com/learn/training/20214>.

Tableau's Big Data

- ▶ To upgrade your Tableau software from the free, 14-day version to free, one-year version, visit <https://www.tableau.com/academic/students>. Be prepared to upload your Asbury student ID to verify your student status.

Homework 3 – Email me (albert.kalim@asbury.edu) your viz link and answers by Sunday, 5/29, 11:59 p.m. ET (10 points total)

Using the data file “xcel_project1_student_data.xls” ([click here](#) to download data)

- ▶ 1. Create a Tableau dashboard with 2 visualizations (viz) based on the data above. You pick the viz types. Then, answer the below questions. (5 points)
- ▶ 2. Which college has the highest enrollment of male students in Fall 2018? (2.5 points)
- ▶ 3. What is the smallest Student ID Anonymous value (i.e., the first value when sorted lowest-to-highest) for a part-time non-resident student enrolled in the Fall 2016 academic term? (2.5 points)